

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



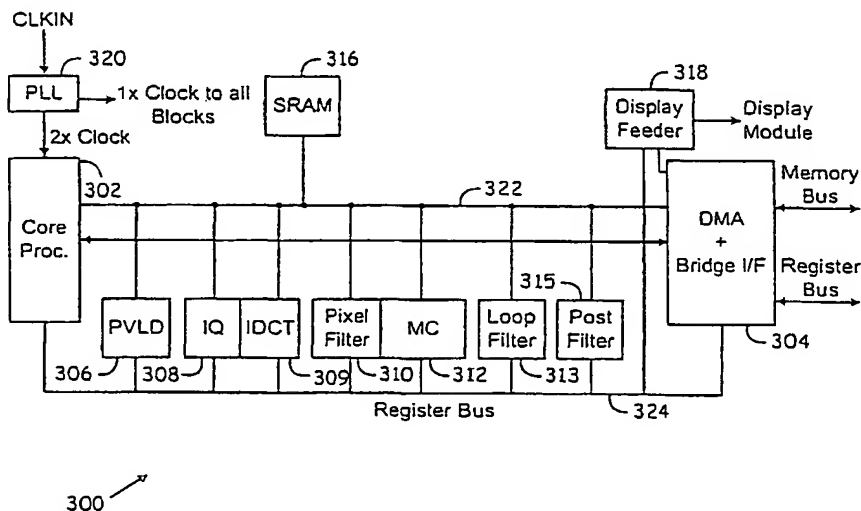
(43) International Publication Date
16 October 2003 (16.10.2003)

PCT

(10) International Publication Number
WO 03/085494 A2

- (51) International Patent Classification⁷: **G06F** 1023 Los Altos Avenue, Los Altos, CA 94022-1017 (US).
XIE, Xiaodong; 39120 Argonaut Way #472, Fremont, CA 94538 (US).
- (21) International Application Number: PCT/US03/09896
- (22) International Filing Date: 1 April 2003 (01.04.2003) (74) Agent: **WIBERG, John, A.**; McAndrews, Held & Malloy, Ltd., 500 West Madison, Suite 3400, Chicago, IL 60661 (US).
- (25) Filing Language: English
- (26) Publication Language: English (84) Designated States (regional): European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR).
- (30) Priority Data: 60/369,144 1 April 2002 (01.04.2002) US Published:
— without international search report and to be republished upon receipt of that report
- (71) Applicant: **BROADCOM CORPORATION** [US/US]; 16215 Alton Parkway, Irvine, CA 92618 (US).
- (72) Inventors: **HSIUN, Vivian**; 5580 Avenida El Cid, Yorba Linda, CA 92887 (US). **MACINNIS, Alexander, G.**; For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: VIDEO DECODING SYSTEM HAVING A PROGRAMMABLE VARIABLE-LENGTH DECODER



(57) Abstract: Video decoding system having a programmable variable-length decoding accelerator. The system includes a decoder processor and a variable-length decoding accelerator. The variable-length decoding accelerator is coupled to the decoder processor and performs variable-length decoding operations on variable-length code in the video data stream. The variable-length decoding accelerator is capable of decoding variable-length code according to any of a plurality of decoding methods. In one embodiment, the variable-length decoder includes a plurality of code tables stored in memory and a code table selection register that is programmable to dictate which of the plurality of code tables is to be utilized to decode variable-length code. In one embodiment, the decoding system includes two variable-length decoding accelerators.

VIDEO DECODING SYSTEM HAVING A PROGRAMMABLE VARIABLE-LENGTH DECODER

PRIORITY CLAIM TO RELATED APPLICATIONS

5 Priority is claimed to U.S. Provisional Patent Application No. 60/369,144, entitled "VIDEO DECODING SYSTEM HAVING A PROGRAMMABLE VARIABLE LENGTH DECODER", filed on April 1, 2002, the subject matter of which is hereby specifically incorporated by reference.

INCORPORATION BY REFERENCE OF RELATED APPLICATIONS

10 The following U.S. Patent Applications are related to the present application and are hereby specifically incorporated by reference: Patent Application No. 10/114,798, entitled "VIDEO DECODING SYSTEM SUPPORTING MULTIPLE STANDARDS"; Patent Application No. 10/114,679, entitled "METHOD OF OPERATING A VIDEO DECODING SYSTEM"; Patent Application No.
15 10/114,797, entitled "METHOD OF COMMUNICATING BETWEEN MODULES IN A DECODING SYSTEM"; Patent Application No. 10/114,886, entitled "MEMORY SYSTEM FOR VIDEO DECODING SYSTEM"; Patent Application No. 10/114,619, entitled "INVERSE DISCRETE COSINE TRANSFORM SUPPORTING MULTIPLE DECODING PROCESSES"; and Patent Application No. 10/113,094,
20 entitled "RISC PROCESSOR SUPPORTING ONE OR MORE UNINTERRUPTIBLE CO-PROCESSORS"; all filed on April 1, 2002; Patent Application No. 10/293,663, entitled "PROGRAMMABLE VARIABLE LENGTH DECODER", filed on November 12, 2002; and Patent Application No. _____, entitled "MEMORY ACCESS ENGINE HAVING MULTI-LEVEL COMMAND
25 STRUCTURE"; and Patent Application No. _____, entitled "INVERSE QUANTIZER SUPPORTING MULTIPLE DECODING PROCESSES"; both filed on April 1, 2003.

FIELD OF THE INVENTION

30 The present invention relates generally to video decoding systems, and, more particularly, to variable-length decoding.

BACKGROUND OF THE INVENTION

Digital video decoders decode compressed digital data that represent video images in order to reconstruct the video images. Most transmitted video data is compressed and decompressed using, among other techniques, variable-length coding, such as Huffman coding. Huffman coding is a widely used technique for lossless data compression that achieves compact data representation by taking advantage of the statistical characteristics of the source. The Huffman code is a prefix-free variable-length code that assures that a code is uniquely decodable. In Huffman code, no codeword is the prefix of any other codeword. In some video compression formats, run-length processed data are often subsequently coded by variable-length coding for further data compression.

Variable-length encoding following the Huffman coding principle allocates codes of different lengths to different input data according to the probability of occurrence of the input data, so that statistically more frequent input codes are allocated shorter codes than the less frequent codes. The less frequent input codes are allocated longer codes. The allocation of codes may be done either statically or adaptively. For the static case, the same output code is provided for a given input datum, no matter what block of data is being processed. For the adaptive case, output codes are assigned to input data based on a statistical analysis of a particular input block or set of blocks of data, and possibly changes from block to block (or from a set of blocks to a set of blocks).

A relatively wide variety of encoding/decoding algorithms and encoding/decoding standards presently exists, and many additional algorithms and standards are sure to be developed in the future. The various algorithms and standards produce compressed video bitstreams of a variety of formats. Some existing public format standards include MPEG-1, MPEG-2 (used for standard definition, or SD, and high definition, or HD), MPEG-4, H.263, H.263+ and MPEG-4 AVC, also called H.264. Also, private standards have been developed by Microsoft Corporation (Windows Media), RealNetworks, Inc., Apple Computer, Inc. (QuickTime), and others. The combination of run-length coding and Huffman coding has been adopted in most compression/decompression standards. However, every

standard has its own variable length code tables and run-length definitions. It would be desirable to have a multi-format decoding system that can decode a variety of variable-length encoded bitstream formats, including existing and future standards, and to do so in a cost-effective manner.

5 A highly optimized hardware architecture can be created to address a specific video decoding standard, but this kind of solution is typically limited to a single format. On the other hand, a fully software based solution is capable of handling any encoding format, but at the expense of performance. Currently, the latter case is solved in the industry by the use of general-purpose processors running on personal
10 computers. Sometimes the general-purpose processor is accompanied by digital signal processor (DSP) oriented acceleration modules, like multiply-accumulate (MAC), that are intimately tied to the particular internal processor architecture. For example, in one existing implementation, an Intel Pentium processor is used in conjunction with an MMX acceleration module. Such a solution is limited in
15 performance and does not lend itself to creating mass market, commercially attractive systems.

 Others in the industry have addressed the problem of accommodating different encoding/decoding algorithms by designing special purpose DSPs in a variety of architectures. Some companies have implemented Very Long Instruction Word
20 (VLIW) architectures more suitable to video processing and able to process several instructions in parallel. In these cases, the processors are difficult to program when compared to a general-purpose processor, and VLIW processors tend to have difficulty decoding variable length codes since the nature of the codes does not lend itself to parallel operations. In special cases, where the processors are dedicated for
25 decoding compressed video, special processing accelerators are tightly coupled to the instruction pipeline and are part of the core of the main processor.

 Yet others in the industry have addressed the problem of accommodating different encoding/decoding algorithms by simply providing multiple instances of hardware dedicated to a single algorithm. This solution is inefficient and is not cost-
30 effective, and it not practical for all compressed video formats. Thus there is a need

for a simple and flexible decoding system that can speedily and efficiently decode variable-length codes of varying standards.

Further limitations and disadvantages of conventional and traditional approaches will become apparent to one of skill in the art through comparison of such systems with the present invention as set forth in the remainder of the present application with reference to the drawings.

SUMMARY OF THE INVENTION

One aspect of the present invention is directed to a video decoding system comprising a decoder processor, a first variable-length decoding accelerator and a second variable-length decoding accelerator. The decoder processor is adapted to perform decoding functions on a video data stream. The first and second variable-length decoding accelerators are each coupled to the decoder processor and are adapted to perform variable-length decoding operations on variable-length code in the video data stream.

Another aspect of the present invention is directed to a variable-length decoder having a plurality of code tables and a code table selection register. The code tables are stored in memory. Each code table corresponds to either a different class of variable length codes in a decoding method or to a different decoding method. Each of the code tables matches variable-length codes to their corresponding decoded information. The code table selection register holds a value that dictates which of the plurality of code tables is to be utilized to decode variable-length code. The register is programmable to dictate the appropriate code table to be employed according to the format of an incoming data stream.

Another aspect of the present invention is directed to a video decoding system having a decoder processor and a variable-length decoding accelerator. The decoder processor performs decoding functions on a video data stream. The variable-length decoding accelerator is coupled to the decoder processor and performs variable-length decoding operations on variable-length codes in the video data stream. The variable-length decoding accelerator is capable of decoding variable-length codes according to any of a plurality of decoding methods.

It is understood that other embodiments of the present invention will become readily apparent to those skilled in the art from the following detailed description, wherein embodiments of the invention are shown and described only by way of illustration of the best modes contemplated for carrying out the invention. As will be realized, the invention is capable of other and different embodiments and its several details are capable of modification in various other respects, all without departing from the spirit and scope of the present invention. Accordingly, the drawings and detailed description are to be regarded as illustrative in nature and not as restrictive.

DESCRIPTION OF THE DRAWINGS

These and other features, aspects, and advantages of the present invention will become better understood with regard to the following description, appended claims, and accompanying drawings where:

FIG. 1 is a functional block diagram of a digital media system in which the present invention may be illustratively employed.

FIG. 2 is a functional block diagram demonstrating a video decode data flow according to an illustrative embodiment of the present invention.

FIG. 3 is a functional block diagram of a decoding system according to an illustrative embodiment of the present invention.

FIG. 4 is a functional block diagram representing a variable-length decoding system according to an illustrative embodiment of the present invention.

FIG. 5 is a flowchart representing a method of variable-length decoding a digital video data stream according to an illustrative embodiment of the present invention.

FIG. 6 is a block diagram showing stream funnel and codeword search engine elements of a variable-length decoder according to an illustrative embodiment of the present invention.

FIG. 7 is a flowchart representing a method of decoding a variable-length code data stream according to an illustrative embodiment of the present invention.

FIG. 8 is an example of a code table according to the code table storage algorithm of the present invention.

FIG. 9 is a chart representing a decoding pipeline according to an illustrative embodiment of the present invention.

5 FIG. 10 is a flowchart representing a macroblock decoding loop according to an illustrative embodiment of the present invention.

FIG. 11 is a functional block diagram of a digital video decoding system according to an illustrative embodiment of the present invention.

10 FIG. 12 is a chart representing a decoding pipeline according to an illustrative embodiment of the present invention.

FIG. 13 is a chart representing a dual-row decoding pipeline employing cycle stealing according to an illustrative embodiment of the present invention.

DETAILED DESCRIPTION

15 The present invention forms an integral part of a complete digital media system and provides flexible decoding resources. FIG. 1 is a functional block diagram of a digital media system in which the present invention may be illustratively employed. It will be noted, however, that the present invention can be employed in systems of widely varying architectures and widely varying designs.

20 The digital media system of FIG. 1 includes transport processor 102, audio decoder 104, direct memory access (DMA) controller 106, system memory controller 108, system memory 110, host CPU interface 112, host CPU 114, digital video decoder 116, display feeder 118, display engine 120, graphics engine 122, display encoders 124 and analog video decoder 126. The transport processor 102 receives and processes a digital media data stream. The transport processor 102 provides the
25 audio portion of the data stream to the audio decoder 104 and provides the video portion of the data stream to the digital video decoder 116. In one embodiment, the audio and video data is stored in main memory 110 prior to being provided to the audio decoder 104 and the digital video decoder 116. The audio decoder 104 receives the audio data stream and produces a decoded audio signal. DMA controller 106

controls data transfer amongst main memory 110 and memory units contained in elements such as the audio decoder 104 and the digital video decoder 116. The system memory controller 108 controls data transfer to and from system memory 110. In an illustrative embodiment, system memory 110 is a dynamic random access memory (DRAM) unit. The digital video decoder 116 receives the video data stream, decodes the video data and provides the decoded data to the display engine 120 via the display feeder 118. The analog video decoder 126 digitizes and decodes an analog video signal (e.g. NTSC or PAL) and provides the decoded data to the display engine 120. The graphics engine 122 processes graphics data in the data stream and provides the processed graphics data to the display engine 120. The display engine 120 prepares decoded video and graphics data for display and provides the data to display encoders 124, which provide an encoded video signal to a display device.

FIG. 2 is a functional block diagram demonstrating a video decode data flow according to an illustrative embodiment of the present invention. Transport streams are parsed by the transport processor 102 and written to main memory 110 along with access index tables. The video decoder 116 retrieves the compressed video data for decoding, and the resulting decoded frames are written back to main memory 110. Decoded frames are accessed by the display feeder interface 118 of the video decoder for proper display by a display unit. In FIG. 2, two video streams are shown flowing to the display engine 120, suggesting that, in an illustrative embodiment, the architecture allows multiple display streams by means of multiple display feeders.

Aspects of the present invention relate to the architecture of digital video decoder 116. In accordance with an exemplary embodiment of the present invention, a moderately capable general purpose CPU with widely available development tools is used to decode a variety of coded streams using hardware accelerators designed as integral parts of the decoding process.

FIG. 3 is a functional block diagram of a digital video decoding system 300 according to an illustrative embodiment of the present invention. The digital video decoding system 300 of FIG. 3 can illustratively be employed to implement the digital video decoder 116 of Figures 1 and 2. Video decoding system 300 includes core decoder processor 302, DMA Bridge 304, decoder memory 316, display feeder 318,

phase-locked loop element 320, and acceleration modules 306, 308, 309, 310, 312, 313 and 315. The acceleration modules include variable-length decoder (VLD) 306, inverse quantization (IQ) module 308, inverse discrete cosine transform (IDCT) module 309, pixel filter 310, motion compensation module 312, loop filter 313 and post filter 315. The acceleration modules 306, 308, 309, 310 and 312 are hardware accelerators that accelerate special decoding tasks that would otherwise be bottlenecks for real-time video decoding if these tasks were handled by the core processor 302, alone. This helps the core processor achieve the required performance.

The core processor 302 is the central control unit of the decoding system 300. The core processor 302 prepares the data for decoding. The core processor 302 also orchestrates the macroblock (MB) processing pipeline for the acceleration modules and fetches the required data from main memory 110 via the DMA bridge 304. The core processor 302 also handles some data processing tasks. Picture level processing, including sequence headers, GOP headers, picture headers, time stamps, macroblock-level information, except the block coefficients, and buffer management, are performed directly and sequentially by the core processor 302, without using the accelerators 304, 306, 308, 309, 310, 312, 313 and 315, other than the VLD 306 (which accelerates general bitstream parsing). Picture level processing does not generally overlap with slice level/macroblock decoding. In an illustrative embodiment of the present invention, the core processor 302 is a MIPS processor, such as a MIPS32 implementation, for example.

The most widely-used compressed video formats fall into a general class of DCT-based, variable-length coded, block-motion-compensated compression algorithms. As mentioned above, these types of algorithms encompass a wide class of international, public and private standards, including MPEG-1, MPEG-2 (SD/HD), MPEG-4, H.263, H.263+, H.264, MPEG-4 AVC, Microsoft Corp., Real Networks, QuickTime, and others. Each of these algorithms implements some or all of the functions implemented by variable-length decoder 306, and the other hardware accelerators 308, 309, 310 312, 313 and 315, in different ways that prevent fixed hardware implementations from addressing all requirements without duplication of resources. In accordance with one aspect of the present invention, variable-length

decoder 306 and the other hardware accelerators 308, 309, 310, 312, 313 and 315, are internally programmable to allow changes according to various processing algorithms. This enables a decoding system that decodes most standards efficiently and flexibly.

5 The decoding system of the present invention employs high-level granularity acceleration with internal programmability to achieve the requirements above by implementation of very fundamental processing structures that can be configured dynamically by the core decoder processor. This contrasts with a system employing fine-granularity acceleration, such as multiply-accumulate (MAC), adders, multipliers, FFT functions, DCT functions, etc. In a fine-granularity acceleration
10 system, the decompression algorithm has to be implemented with firmware that uses individual low-level instructions (like MAC) to implement a high-level function, and each instruction runs on the core processor. In the high-level granularity system of the present invention, the firmware configures, i.e., programs, variable-length decoder 306 and the other hardware accelerators 308, 309, 310, 312, 313 and 315, which in
15 turn represent high-level functions (like variable-length decoding) that run without intervention from the main core processor 302. Therefore, each hardware accelerator 306, 308, 309, 310, 312, 313 and 315, runs in parallel according to a processing pipeline dictated by the firmware in the core processor 302. Upon completion of the high-level functions, each accelerator notifies the main core processor 302, which in
20 turn decides what the next processing pipeline step should be.

In an illustrative embodiment of the present invention, the software control consists of a simple pipeline that orchestrates decoding by issuing commands to each hardware accelerator module for each pipeline stage, and a status request mechanism that makes sure that all modules have completed their pipeline tasks before issuing the
25 start of the next pipeline stage. As used in the present application, the term "stage" can refer to all of the decoding functions performed during a given time slot, or it can refer to a functional step, or group of functional steps, in the decoding process. Each hardware module 306, 308, 309, 310, 312, 313 and 315, performs its task after being so instructed by the core processor 302. In an illustrative embodiment of the present
30 invention, each hardware module includes a status register that indicates whether the module is active or inactive. The core processor 302 polls the status register to

determine whether the hardware module has completed its task. In an alternative embodiment, the hardware accelerators share a status register.

Variable-length decoder 306 is a hardware accelerator that accelerates the process of decoding variable-length codes, which might otherwise be a bottleneck for a decoding process if it were handled by the core processor 302 alone. In accordance with the present invention, the VLD 306 can also be any other type of entropy decoder. But for purposes of explanation, the present invention will be described with respect to a variable-length decoder. The VLD 306 performs decoding of variable length codes (VLC) in the compressed bit stream to extract coefficients, such as DCT coefficients, from the compressed data stream. Different coding formats generally have their own special VLC tables. According to the present invention, the VLD module 306 is internally programmable to allow changes according to various processing algorithms. The VLD 306 is completely configurable in terms of the VLC tables it can process. The VLD 306 can accommodate different VLC tables, selectable as needed under the control of the core processor 302. In an illustrative embodiment of the present invention, the VLD 306 includes a register that the core processor can program to guide the VLD 306 to use the appropriate VLC table according to the needs of the encoding/decoding algorithm and the class of codes expected.

The VLD 306 is designed to support the worst-case requirement for VLD operation, such as with MPEG-2 HDTV (Main Profile at High Level) for video decoding, while retaining its full programmability. The VLD 306 includes a code table random access memory (RAM) for fastest performance. Some compression/decompression formats, such as Windows Media Technology 8 (WMT8) video, may require larger code tables that do not fit entirely within the code RAM in the VLD 306. For such cases, according to an illustrative embodiment of the present invention, the VLD 306 can make use of both the decoder memory 316 and the main memory 110 as needed. Performance of VLC decoding is reduced somewhat when codes are searched in video memory 316 and main memory 110. Therefore, for formats that require large amounts of code, the most common codes are stored in the VLD code RAM, the next most common codes are stored in decoder memory 316,

and the least common codes are stored in main memory 110. Also, such codes are stored in decoder memory 316 and main memory 110 such that even when extended look-ups in decoder memory 316 and main memory 110 are required, the most commonly occurring codes are found more quickly. This allows the overall performance to remain exceptionally high. The VLD 306 decodes variable length codes in as little as one clock, depending on the specific code table in use and the specific code being decoded.

In an illustrative embodiment of the present invention, the VLD 306 helps the core processor 104 to decode header information in the compressed bitstream. In an illustrative embodiment of the present invention, the VLD module 306 is architected as a coprocessor to the decoder processor 110. That is, it can operate on a single-command basis where the core processor issues a command (via a coprocessor instruction) and waits (via a Move From Coprocessor instruction) until it is executed by the VLD 306, without polling to determine completion of the command. This increases performance when a large number of VLC codes that are not DCT coefficients are parsed.

In an alternative embodiment, the VLD 306 is architected as a hardware accelerator. In this embodiment, the VLD 306 can perform complex tasks such as decoding a set of VLC codes, and it includes a status register that indicates whether the module is active or inactive. The core processor 302 polls the status register to determine whether the VLD 306 has completed its tasks. In an alternative embodiment, the VLD 306 shares a status register with other decoding elements, such as decoding elements 308, 309, 310 and 312.

In an illustrative embodiment of the present invention, the VLD module 306 includes two variable-length decoders. Each of the two variable-length decoders can be hardwired to efficiently perform decoding according to a particular compression standard, such as MPEG-2 HD. In an illustrative embodiment, one or both of two VLDs can be optionally set as a programmable VLD engine, with a code RAM to hold VLC tables for other media coding formats. The two VLD engines are controlled independently by the core processor 302, and either one or both of them will be employed at any given time, depending on the application.

5 The VLD 306 can operate on a block-command basis where the core processor 302 commands the VLD 306 to decode a complete block of VLC codes, such as DCT coefficients, and the core processor 302 continues to perform other tasks in parallel. In this case, the core processor 302 verifies the completion of the block operation by checking a status bit in the VLD 306. The VLD 306 produces results (tokens) that are stored in decoder memory 316.

The VLD 306 checks for invalid codes and recovers gracefully from them. Invalid codes may occur in the coded bit stream for a variety of reasons, including errors in the video encoding, errors in transmission, and discontinuities in the stream.

10 The inverse quantizer module 308 performs run-level code (RLC) decoding, inverse scanning (also called zig-zag scanning), inverse quantization and mismatch control. The coefficients, such as DCT coefficients, extracted by the VLD 306 are processed by the inverse quantizer 308 to bring the coefficients from the quantized domain to the DCT domain. In an exemplary embodiment of the present invention, 15 the IQ module 308 obtains its input data (run-level values) from the decoder memory 316, as the result of the VLD module 306 decoding operation. In an alternative embodiment, the IQ module 308 obtains its input data directly from the VLD 306. This alternative embodiment is illustratively employed in conjunction with encoding/decoding algorithms that are relatively more involved, such as MPEG-2 HD 20 decoding, for best performance. The run-length, value and end-of-block codes read by the IQ module 308 are compatible with the format created by the VLD module when it decodes blocks of coefficient VLCs, and this format is not dependent on the specific video coding format being decoded.

25 The IDCT module 309 performs the inverse transform to convert the coefficients produced by the IQ module 308 from the frequency domain to the spatial domain. The primary transform supported is the discrete cosine transform (DCT) as specified in MPEG-2, MPEG-4, IEEE, and several other standards. The IDCT module 309 also supports alternative related transforms, such as the "linear" transform in H.264 and MPEG-4 AVC, which is not quite the same as IDCT.

30 In an illustrative embodiment of the present invention, the coefficient input to the IDCT module 309 is read from decoder memory 316, where it was placed after

inverse quantization by the IQ module 308. The transform result is written back to decoder memory 316. In an exemplary embodiment, the IDCT module uses the same memory location in decoder memory 316 for both its input and output, allowing a savings in on-chip memory usage. In an alternative embodiment, the coefficients produced by the IQ module are provided directly to the IDCT module 309, without first depositing them in decoder memory 316. To accommodate this direct transfer of coefficients, in one embodiment of the present invention, the IQ module 308 and IDCT module 309 are part of the same hardware module and use a common interface to the core processor. In an exemplary embodiment, the transfer of coefficients from the IQ module 308 to the IDCT module 309 can be either direct or via decoder memory 316. For encoding/decoding algorithms that are relatively more involved, such as MPEG-2 HD decoding, the transfer is direct in order to save time and improve performance.

The pixel filter 310 performs pixel filtering and interpolation as part of the motion compensation process. Motion compensation is performed when an image from a region of a previous frame is similar to a region in the present frame, just at a different location within the frame. Rather than recreate the image anew from scratch, the previous image is used and just moved to the proper location within the frame. For example, assume the image of a person's eye is contained in a macroblock of data at frame #0. Say that the person moved to the right so that at frame #1 the same eye is located in a different location in the frame. Motion compensation uses the eye from frame #0 (the reference frame) and simply moves it to the new location in order to get the new image. The new location is indicated by motion vectors that denote the spatial displacement in frame #1 with respect to reference frame #0.

The pixel filter 310 performs the interpolation necessary when a reference block is translated (motion-compensated) into a position that does not land on whole-pixel locations. For example, a hypothetical motion vector may indicate to move a particular block 10.5 pixels to the right and 20.25 pixels down for the motion-compensated prediction. In an illustrative embodiment of the present invention, the motion vectors are decoded by the VLD 306 in a previous processing pipeline stage and are stored in the core processor 302. Thus, the pixel filter 310 gets the motion

information as vectors and not just bits from the bitstream during decoding of the “current” macroblock in the “current” pipeline stage. The reference block data for a given macroblock is stored in memory after decoding of said macroblock is complete. In an illustrative embodiment, the reference picture data is stored in system memory 110. If and when that reference macroblock data is needed for motion compensation of another macroblock, the pixel filter 310 retrieves the reference macroblock pixel information from system memory 110 and the motion vector from the core processor 302 and performs pixel filtering. The pixel filter stores the filtered result (pixel prediction data) in decoder memory 316.

The motion compensation module 312 reconstructs the macroblock being decoded by performing the addition of the decoded difference (or “error”) pixel information from the IDCT 309 to the pixel prediction data from the output of the pixel filter 310. The pixel filter 310 and motion compensation module 312 are shown as one module in FIG. 3 to emphasize a certain degree of direct cooperation between them.

The loop filter 313 and post filter 315 perform de-blocking filter operations. Some decoding algorithms employ a loop filter and others employ a post filter. The difference is where in the processing pipeline each filter 313, 315 does its work. The loop filter 313 processes data within the reconstruction loop and the results of the filter are used in the actual reconstruction of the data. The post filter 315 processes data that has already been reconstructed and is fully decoded in the two-dimensional picture domain. In an illustrative embodiment of the present invention, the loop filter 313 and post filter 315 are combined in one filter module.

In an illustrative embodiment the input data to the loop filter 313 and post filter 315 comes from decoder memory 316. This data includes pixel and block/macroblock parameter data generated by other modules in the decoding system 300. In an illustrative embodiment of the present invention, the loop filter 313 and post filter 315 have no direct interfaces to other processing modules in the decoding system 300. The output data from the loop filter 313 and post filter 315 is written into decoder memory 316. The core processor 302 then causes the processed data to be put in its correct location in main memory.

At the macroblock level, the core processor 302 interprets the decoded bits for the appropriate headers and decides and coordinates the actions of the hardware blocks 306, 308, 309, 310, 312, 313 and 315. Specifically, all macroblock header information, from the macroblock address increment (MBAinc) to motion vectors (MVs) and to the cbp pattern, in the case of MPEG-2 decoding, for example, is derived by the core processor 302. The core processor 302 stores related information in a particular format or data structure (determined by the hardware module specifications) in the appropriate buffers in the decoder memory 316. For example, the quantization scale is passed to the buffer for the IQ engine 308; macroblock type, motion type and pixel precision are stored in the parameter buffer for the pixel filter engine 310. The core processor keeps track of certain information in order to maintain the correct pipeline. For example, motion vectors of the macroblock are kept as the predictors for future motion vector derivation.

Decoder memory 316 is used to store macroblock data and other time-critical data used during the decode process. Each hardware block 306, 308, 309, 310, 312, 314 accesses decoder memory 316 to either read the data to be processed or write processed data back. In an illustrative embodiment of the present invention, all currently used data is stored in decoder memory 316 to minimize access to main memory. Each hardware module 306, 308, 309, 310, 312, 314 is assigned one or more buffers in decoder memory 316 for data processing. Each module accesses the data in decoder memory 316 as the macroblocks are processed through the system. In an exemplary embodiment, decoder memory 316 also includes parameter buffers that are adapted to hold parameters that are needed by the hardware modules to do their job at a later macroblock pipeline stage. The buffer addresses are passed to the hardware modules by the core processor 302. In an illustrative embodiment, decoder memory 316 is a static random access memory (SRAM) unit.

The core processor 302, DMA Bridge 304, VLD 306, IQ 308, IDCT 309, pixel filter 310, motion compensation module 312, loop filter 313 and post filter 315 have access to decoder memory 316 via the internal bus 322. The VLD 306, IQ 308, IDCT 309, pixel filter 310, motion compensation module 312, loop filter 313 and post filter 315 use the decoder memory 316 as the source and destination memory for their

normal operation. The CPU 114 has access to decoder memory 316, and the DMA engine 304 can transfer data between decoder memory 316 and the main system memory (DRAM) 110. The arbiter for decoder memory 316 is in the bridge module 304.

5 The bridge module 304 arbitrates and moves picture data between decoder memory 316 and main memory. The bridge interface 304 includes an internal bus network that includes arbiters and a direct memory access (DMA) engine. The DMA bridge 304 serves as an asynchronous interface to the system buses.

10 The display feeder module 318 reads decoded frames from main memory and manages the horizontal scaling and displaying of picture data. The display feeder 318 interfaces directly to a display module. In an illustrative embodiment, the display feeder 318 includes multiple feeder interfaces, each including its own independent color space converter and scaler. The display feeder 318 handles its own memory requests via the bridge module 304.

15 In an illustrative embodiment of the present invention, the core processor 302 runs at twice the frequency of the other processing modules 306, 308, 309, 310, 312, 313, 315. An elegant, flexible and efficient clock strategy is achieved by generating two internal clocks in an exact 2:1 relationship to each other. The system clock signal CLKIN is used as input to the phase-locked loop element (PLL) 320, which is a
20 closed-loop feedback control system that locks to a particular phase of the system clock to produce a stable signal with little jitter. The PLL element 320 generates a 1X clock for the hardware accelerators, DMA bridge 304 and the core processor bus interface, while generating a 2X clock for the core processor 302 and the core processor bus interface. This is to allow the core processor 302 to operate at high
25 clock frequencies if it is designed to do so, and to allow other logic to operate at the slower 1x clock frequency. It also allows the decoding system 300 to run faster than the nominal clock frequency if the circuit timing supports it.

30 FIG. 4 is a functional block diagram representing a variable-length decoding system according to an illustrative embodiment of the present invention. In FIG. 4, elements that are also shown in FIG. 3 are given like reference numbers. The VLD 306 includes decoder processor interface 400, stream funnel 402, codeword search

engine 404, block buffer 406, decoder memory interface 408, code table selection register 412 and status register 414.

The input 410 to the VLD 306 is a bit stream without explicit word boundaries. The VLD 306 decodes a codeword, determines its length, and shifts the input data stream by the number of bits corresponding to the decoded code length, before decoding the next codeword. These are recursive operations that are not pipelined.

The VLD 306 is implemented based on a small, local, code table memory unit, located in codeword search engine 404, that stores programmable variable length code tables. In an illustrative embodiment, the local memory unit is a random access memory (RAM) unit. A small code table memory unit is achieved by employing a multistage search structure that reduces the storage requirement, enables fast bit extraction and efficiently handles the case of a large number of code tables.

The stream funnel 402 receives data from the source (or coded data buffer) and shifts the data according to the previously decoded code length, so as to output the correct window of bits for the symbols that are being currently decoded. In an illustrative embodiment, the stream funnel receives the incoming bitstream 410 from system memory 110.

The codeword search engine 404 mainly behaves as a symbol search engine. The codeword search engine is based on a multistage search structure. Since codewords are usually assigned based on the probability of appearance, the shortest codeword is generally assigned to the most frequent appearance. The multistage search structure is based on this concept. The codeword search engine 404 incorporates a small code memory that is employed for performing pattern matching. A multistage, pipelined structure is employed to handle the case of a long codeword. Additionally, a code table reduction algorithm can further reduce the storage requirement for a large number of code tables.

Status register 414 is adapted to hold an indicator of the status of the VLD 306. The status register is accessible by the core decoder processor 302 to determine the status of VLD 306. In an illustrative embodiment, the status register 414 indicates

whether or not the VLD has completed its variable-length decoding functions on the current macroblock. In an alternative embodiment of the present invention, the VLD module 306 is architected as a coprocessor to the decoder processor 302. That is, the VLD 306 can operate on a single-command basis where the core processor issues a command (via a coprocessor instruction) and waits (via a Move From Coprocessor instruction) until it is executed by the VLD 306, without polling the status register 414 to determine completion of the command.

Code table selection register 412 is adapted to hold a value that dictates which of a plurality of VLD code tables is to be utilized to decode variable-length code. In an illustrative embodiment, code table selection register 412 holds the starting address of the code table to be employed. The code table selection register 412 is programmable to dictate the appropriate code table to be employed according to the format of an incoming data stream and the class of variable length codes that are expected next. In an illustrative embodiment, the core video processor 302 provides a value (an address, for example) to register 412 to point to the code table that is appropriate for the current data stream and the state of decoding the current stream. The code tables can be switched on a syntax element basis, a macroblock-to-macroblock basis or more or less frequently, as required by the application.

FIG. 5 is a flowchart representing a method of variable-length decoding a digital video data stream according to an illustrative embodiment of the present invention. At step 500, video data of a first encoding/decoding format is received. At step 510, variable-length decoder 306 is configured based on the first encoding/decoding format. In an illustrative embodiment, the core video processor 302 configures variable-length decoder 306 by programming code table selection register 412. At step 520, video data of the first encoding/decoding format is decoded by the variable-length decoder 306. At step 530, video data of a second encoding/decoding format is received. At step 540, the variable-length decoder 306 is configured based on the second encoding/decoding format. Then, at step 550, video data of the second encoding/decoding format is decoded using the variable-length decoder 306. As is described above with respect to the individual hardware

accelerators of Figures 3 and 4, the programming for different decoding formats is done through register bus read and write.

FIG. 6 is a block diagram showing the stream funnel 402 and codeword search engine 404 elements of VLD 306, according to an illustrative embodiment of the present invention.

Stream funnel 402 includes data stream input buffer 600, register D₀ 602, register D₁ 604, left-shifter 606, register D₂ 608, and accumulator 610. The input data (coded stream) are stored in input buffer 600, which, in an illustrative embodiment, is a first-in first-out (FIFO) buffer. The input buffer 600 provides the data to register D₀ 602. Register D₁ 604 also stores part of the incoming bitstream by virtue of load operations that will be discussed below and which load data from register D₀ 602 into register D₁. The contents of registers D₀ and D₁ are in turn provided to left shifter 606. In an illustrative embodiment of the present invention, registers D₀ and D₁ comprise a number of bits equal to the maximum code length. In an embodiment wherein the maximum code length is less than or equal to 32 bits (such as in most video decoding standards), registers D₀ 602 and D₁ 604 each are 32-bit registers, and left-shifter 606 can hold up to 64 bits. Register D₂ 608 indicates the number of bits in register D₁ 604 for which the codeword search engine 604 most recently performed a codeword search. If registers D₀ and D₁ each hold 32 bits, the number of bits indicated by register D₂ can lie between 0 and 31. This number controls the left shifter 606. After the codeword search engine 404 performs a codeword search for a group of bits in register D₁, register D₂ indicates the number of bits just searched. Left shifter 606 then shifts the indicated number of bits to the left so that the first unsearched bit appears at the most significant bit of the output of the left shifter 606.

Accumulator 610 accumulates the number of bits in register D₁ 604 that have been searched by codeword search engine 404 over multiple codeword searches. When the accumulated code length (the number of bits that have been searched) is greater than or equal to the size of register D₁ 604 (for example, 32 bits), a carry-out bit 612 becomes 1. This indicates that all the bits in register D₁ 604 have been used and that register D₀ might not contain the whole next codeword. In that case, a "load" signal is generated. When the "load" signal is generated, the contents of register D₀

602 are loaded into register D₁ 604, a new data word (32 bits in the illustrative example) from the input buffer 600 is loaded into D₀, and the left shifter 606 shifts by the number of bits indicated by register D₂ 608 to the new position, all at substantially the same time, to prepare for the next search/decode cycle. If the accumulated code length is not greater than or equal to the size of register D₁ 604 (e.g., 32), the carry-out signal 612 is 0. Assuming the maximum code length is 32 bits (the size of registers D₀ 602 and D₁ 604 in the illustrative embodiment), since at least 32 bits of data in register D₀ 602 and D₁ 604 are not used yet, there are always enough bits for the next search/decoding cycle. Registers D₀ 602 and D₁ 604 remained unchanged if the accumulated code length is not greater than or equal to the size of the registers D₀ 602 and D₁ 604.

When the accumulated code length is greater than or equal to the size of registers D₀ and D₁, and there is no data available in the input buffer 600, the decoding pipes are put on hold. In other words, the contents of register D₀ 602 are not loaded into register D₁ 604. The decoding processing then waits until data is available in the input buffer 600.

Codeword search engine 404 includes an address generator 612 and a local memory unit 514. Address generator 612 generates a memory address at which to perform a codeword search. In an illustrative embodiment, this address will reside in the local memory unit 614, but it may also reside in decoder memory 316 or system memory 110, as will be described below. The address generator 612 generates the address to be searched by adding the value of the bits retrieved from left shifter 606, i.e., the data for which a search is to be performed, to a base address. For the first search performed in a given code table, and for subsequent searches when the previous search yielded a code match, the base address is equal to the start address of the code table to be searched. For subsequent searches performed after a previous search did not yield a code match, the base address is equal to the sum of the start address of the code table plus an offset that was indicated by the code table entry of the previous search.

In an illustrative embodiment of the present invention, the starting address of the code table to be searched can be programmed. In this way, the appropriate code

table can be selected for the encoding/decoding format and the current state of the bitstream being decoded. In an illustrative embodiment of the present invention, code table selection register 412 holds the starting address of the code table to be searched. This register can be accessed by the decoder processor 302 to point to the code table that is appropriate for the current data stream. The code tables can be switched on a syntax element basis, a macroblock-to-macroblock basis or on other intervals.

Local code table memory 614 holds the code look-up table that is to be used during the variable-length decoding process. The code table that starts at the indicated start address is used in decoding the incoming bitstream. In an illustrative embodiment of the present invention, code table memory 614 is a random access memory (RAM) unit. In a further illustrative embodiment, the code table memory 614 is a relatively small memory unit, for example, a 512x32 single-port RAM.

In an illustrative embodiment of the present invention, if a given code look-up table does not fit within the code table memory unit 614, portions of the table can be stored in decoder memory 316 and/or system memory 110. In an illustrative embodiment, if more memory is needed than the local memory unit 614 alone, first the decoder memory 316 is utilized, and if more still is needed, the system memory 110 is utilized. Where multiple memory units are utilized, the shortest, and therefore most common codes, are stored in local code table memory 614. The next-shortest codes are stored in decoder memory 316, and if needed, the longest codes are stored in system memory 110. This architecture allows for fast bit extraction.

According to an illustrative embodiment of the present invention, the codeword search engine 404 employs a code table storage and look-up method that enables fast bit extraction and also reduces the size of the code tables. Reducing the size of the code tables further reduces the storage requirement for a large number of code tables and results in fast performance for a broad range of codes. One embodiment of the code table storage and look-up method makes use of the multiple memory unit structure mentioned above and uses a multistage, pipelined structure to handle the case of a long codeword.

The code table memory unit 614 supports multiple code tables (up to 32 in an illustrative embodiment). In an illustrative embodiment each code table has the

following general information which is pre-programmed by the decoder processor 302: the starting address, in the local memory 614, of the code table during the first search (FSA), the searching length during the first level search (FSL), an indication of whether a sign bit follows the code to be searched, the size of a run code that may follow a specific variable length code (such as an escape code), the size of a level code that may follow a specific variable length code (such as an escape code), and an indication of whether an end of block (EOB) code is to follow the run-level code that may appear. A high sign bit indicator indicates that the code table has a sign bit following the codeword. The size of the run code indicates how many bits are allocated to the run portion of the results. The size of the level code indicates how many bits are allocated to the level portion of the results. The EOB bit indicates whether a "last" bit or EOB bit is expected after the run-level code that may appear. The run-level code and EOB bit may appear following a designated variable length code such as escape (ESC) code. For example, in MPEG4 video, if the escape code is type4, the 15 bits following ESC are decoded as fixed length codes represented by 1-bit LAST, 6-bit RUN and 8-bit LEVEL. The meanings of run, level and EOB can vary between different video coding/decoding formats.

Each address of a code table comprises a code table entry. Each entry includes a current code length (CCL) indicator, a next search length (NSL) indicator, an end-of-block (EOB or last code) bit, a status indicator and an information/offset value. The status indicator indicates whether that entry represents a codeword match. If the entry does represent a codeword match, the information/offset value is the matching information, that is, the data that the just-matched codeword represents (the "meaning" of the codeword). If the entry does not represent a codeword match, the information/offset value indicates an address at which to perform the next codeword search. In an illustrative embodiment of the present invention, the offset value indicates an address at which to base the next search to complete decoding of the current codeword. In an alternative embodiment, the offset value is added to another address to obtain the base address from which to perform the next search to complete decoding of the current codeword.

The status indicator can also indicate other aspects of the search status. For example, if the entry does not represent a codeword match, the status indicator indicates the memory unit in which to perform the next codeword search. Also, if the entry represents an error, i.e., no valid code would result in the entry at that memory location to be reached, the status indicator indicates as much. In an illustrative embodiment of the present invention, the status indicator is a 4-bit word having the meanings shown in Table 1.

Stat us Code [3:0]	Meaning
1 000	Escape code followed by run length code
0 001	Special Codeword 1
1 001	Special Codeword 2
0 010	CodeWord Found
1 010	Goto Next Level Code Search @ Code RAM
0 011	Goto Next Level Code Search @ Decoder Memory
1 011	Error has been detected
0 100	Goto Next Level Code Search @ System Memory
rs othe	reserved

Table 1

As mentioned above, if the code table entry represents a codeword match (status = 0100), the information/offset value represents the meaning of the codeword. If the code table entry does not represent a codeword match, and the next search is to be performed in local memory (status = 0101), the information/offset value represents the start address of the next search level at local memory 614 (code RAM). If the code table entry does not represent a codeword match, and the next search is to be performed in decoder memory (status = 0110), the information/offset value represents

the offset of the secondary code table at the decoder memory 316. If the entry does not represent a codeword match, and the next search is to be performed in system memory (status = 1000), the information/offset value represents the offset of the tertiary code table at the system memory 110.

5 In an illustrative embodiment, the current code-length indicator indicates the number of bits that the input bitstream should be shifted prior to the next codeword search. If the code table entry represents a codeword match, the current code-length represents the number of bits, out of the currently searched group of bits, that correspond to the matched information represented by the information/offset value. If
10 the code table entry does not represent a codeword match, the current code-length indicator indicates the number of bits that were consumed in the current stage of search. If the entry represents an error, the current code-length indicator indicates that no bits in the current search have been matched. In an alternative embodiment, the current code-length indicator indicates the number of bits consumed from the input bit
15 stream when the status indicates a match, and it indicates the number of bits to be searched in the next stage of search when the status indicates no match in the current stage. In such an embodiment the number of bits consumed when there is not a match is implied to be the number of bits searched in the current stage.

 In an illustrative embodiment, each code table entry that does not represent a
20 codeword match further includes a next-search-length (NSL) indicator that indicates the number of bits to perform a codeword search for in the next stage. In such an embodiment, the code table entries that do represent a codeword match do not contain a next-search-length indicator, as the search length in the next stage automatically reverts to an initial value. In an alternative embodiment, the code table entries that do
25 represent a codeword match do contain a next-search-length indicator, which indicates the initial value.

 In an illustrative embodiment, the end-of-block bit is high if the just-decoded code is the last code in a block of codes to be decoded. In alternative embodiments the end-of-block bit has the opposite polarity, or the end-of-block bit is not included.

30 The code table memory 614 and the address generator 612 work together to perform pattern matching on the data stream. When a codeword is matched at a code

table entry, the status indicator in the entry will indicate that that is the case. If an accessed code table entry is not a match, the state machine will go to the next stage to keep searching until the codeword is found. If the status indicator shows that an error has occurred, the VLD 306 will stop searching the next codeword, set an error status bit to "1," report the error to the decoder processor 302 and enter an idle state.

FIG. 7 is a flowchart representing a method of decoding a variable-length code data stream according to an illustrative embodiment of the present invention. The method implements a code table storage algorithm, and a method of traversing a code table implementing the algorithm, that reduces the storage requirement and enables fast code look-up. At step 705, the appropriate code table is loaded according to the compression/decompression standard of the data stream being decoded. The code table is illustratively loaded into local memory 614. The start address of the code table in the local memory is designated m . At step 710, a base memory address is set equal to the start address. Also at step 710, the search length, n , i.e., the number of bits from the data stream for which a code match is sought in a given search, is initialized as a first search length (FSL) value.

At step 715, the next n bits in the data stream are retrieved. In an illustrative embodiment, these bits are retrieved from the n most significant bits of left shifter 606. At step 720, the address at which to search for a code match is generated by adding the value of the n bits retrieved from the bitstream to the base address. This step is illustratively performed by address generator 612. At step 725, the memory location having the address generated in step 720 is accessed, and the status indicator at that memory location is examined. Decision box 730 asks whether the status indicator indicates that a codeword match is found. If the answer is yes, the corresponding information (decoded data), indicated by the information/offset value of the memory location, is output. In an illustrative embodiment of the present invention, the decoded data comprises run and level information which implies the value of one or more transform coefficients, such as discrete cosine transform (DCT) coefficients. In alternative embodiments, the decoded data comprises transform coefficients directly, or other information as appropriate to the video compression/decompression standard being decoded.

If the status indicator indicates that a codeword match is not found, decision box 740 asks whether the status indicator indicates that an error has occurred. Such an error would arise, for example, if the memory location arrived at does not correspond to a valid code. If there is an error, an error indication is given, as indicated at step 745. If the status indicator indicates that either a codeword match is found or an error has occurred, the base address is set equal to the start address, as indicated by step 755, and the search length, n , is set equal to the first search length (FSL), as shown at step 760. If the status indicator indicates that the memory location does not represent a codeword match, and an error has not occurred, the base address is set according to the offset value indicated by the information/offset value, as indicated at step 750, and the search length, n , is set equal to the next-search-length value held in the memory location. In an illustrative embodiment, the search length remains constant throughout the decoding process. In that case, steps 760 and 765 of FIG. 5 can be eliminated.

At step 770, the incoming bitstream is shifted by an amount indicated by the current code-length indicator of the memory location. Step 770 is illustratively performed by left shifter 606. In an illustrative embodiment, if the memory location represents a codeword match, the current code-length indicator indicates the number of the input bits consumed by the most recent search stage in decoding the current code word.. In a further illustrative embodiment, if the memory location represents a non-match, the value of the current code-length indicator is equal to n bits (the number of bits for which the current search was performed). In another embodiment, if the status indicator indicates an error, the value of the current code-length indicator is zero. After step 770, the next n bits in the data stream are accessed, as indicated by step 715, and the above-described process is repeated starting at that point. In an exemplary embodiment, this process is iteratively repeated as long as there is data in the data stream to decode.

FIG. 8 is an example of a code table according to the code table storage algorithm of the present invention. In an illustrative embodiment of the present invention, the code table of FIG. 8 is stored in local memory 614. The following codebook (Table 2) is used in the exemplary code table of FIG. 8:

Codeword	Code Length	Decoded Symbol
1	1	A
010	3	B
011	3	C
.....		
00010	5	X
000110	6	Y
000111	6	Z

Table 2

Each of the addresses 800 in the code table of FIG. 8 represents a codebook entry. The other columns 810, 820, 830, 840 and 850 represent elements of each codebook entry. These elements include current code-length indicator 810, next-search-length indicator 820, end-of-block bit 830, status indicator 840 and information/offset value 850. The illustrative code table of FIG. 8 has a first search length (FSL) of 3 and a starting address (FSA) of 0.

To demonstrate how the code table of FIG. 8 is structured and to demonstrate how it is traversed in order to decode a variable-length bitstream, assume the bits in the most-significant position of left shifter 606 are the bits 1010 (which we know, from referring to the codebook of Table 2, represent symbols A and B). The codeword search engine decodes these bits as follows. Because the first search length is 3, the first three bits of the data stream (101) are pulled from the data stream, that is, from the left shifter 606. The address generator 612 adds the value of these bits (5) to the starting address (0) to get a search address of 5. The code table entry at address 5 has a status indicator = 0100, which indicates that the entry represents a codeword match (see table 1). Therefore, the information/offset value (A) of the entry is outputted as a decoded value. In an illustrative embodiment of the present invention, the decoded data comprises transform coefficients, such as discrete cosine transform (DCT) coefficients. In an illustrative embodiment, this decoded value is provided to decoder memory 316 and stored there. The current-code-length indicator of the entry

at address 5 is a 1. This value is provided to accumulator 610 and register D₂ 608, indicating that one bit (the first 1, corresponding to the outputted A) was decoded in this stage.

Therefore, in the next stage, prior to performing the next search, the left shifter 606 shifts its contents one bit, putting the bits 010 at the three most-significant positions of left shifter 606. The search length is three (the first search length) because the previous search resulted in a codeword match. Thus, the bits 010 are provided to the address generator 612, which adds the value of these bits (2) to the starting address (0) to get a search address of 2 (the starting address is used as the base address because the previous search yielded a match). The code table entry at address 2 has a status indicator = 0100, which indicates that the entry represents a codeword match. Therefore the information/offset value (B) is outputted as a decoded value. Hence, the input string 1010 has been decoded as AB. The current-code-length indicator of the entry at address 2 is a 3. This value is provided to accumulator 610 and register D₂ 608, indicating that three bits (010, corresponding to the outputted B) were decoded in this stage.

Say, for example, the next bits in the data stream (after the bits 1010) are 00010010 (which represent symbols X and B). Because the value stored in register D₂ 608 from the previous search is 3, prior to performing the next search, the left shifter 606 shifts its contents three bits, putting the bits 000 at the three most-significant positions of left shifter 606. The search length is three (the first search length) because the previous search resulted in a codeword match. Thus, the bits 000 are provided to the address generator 612, which adds the value of these bits (0) to the starting address (0) to get a search address of 0 (the starting address is used as the base address because the previous search yielded a match). The code table entry at address 0 has a status indicator = 0101, which indicates that the entry does not represent a codeword match. Therefore, the information/offset value (8) is provided to address generator 612 to be used in calculating the base address of the next search. The code table entry at address 0 has a next search-length indicator of 3. This value is provided to address generator 612 to indicate the number of bits to be retrieved from the left shifter 606 for the next search. The current-code-length indicator of the entry

at address 0 is a 3. This value is provided to accumulator 310 and register D₂ 608, indicating that the left shifter 606 should shift its contents three bits prior to the next codeword search.

Shifting the contents of left shifter 606 by the indicated three bits puts the bits 100 at the three most-significant positions of left shifter 606. The search length is three, as indicated to the address generator 612 by the next-search-length indicator from the previous stage. Thus, the bits 100 are provided to the address generator 612, which adds the value of these bits (4) to the base address to get the search address. The base address is equal to the start address (0) plus the offset value (8) indicated by the information/offset value from the previous stage. Thus the search address = 0 + 8 + 4 = 12. The code table entry at address 12 has a status indicator = 0100, which indicates that the entry represents a codeword match. Therefore the information/offset value (X) is outputted as a decoded value. The current-code-length indicator of the entry at address 12 is a 2. This value is provided to accumulator 610 and register D₂ 608, indicating that two bits (10, which are the first two bits of the just-searched bits and which are also the last two bits of the just-decoded codeword) were decoded in this stage.

Therefore, in the next stage, prior to performing the next search, the left shifter 606 shifts its contents two bits, putting the bits 010 at the three most-significant positions of left shifter 606. The search length is three (the first search length) because the previous search resulted in a codeword match. Thus, the bits 010 are provided to the address generator 612, which adds the value of these bits (2) to the starting address (0) to get a search address of 2 (the starting address is used as the base address because the previous search yielded a match). The symbol B is decoded at the code table entry at address 2, as was described above.

In an illustrative embodiment of the present invention, multiple memory units are used to store the codeword look-up table. For example, in one embodiment, part of the codeword look-up table is stored in local memory 614, part is stored in decoder memory 316, and part is stored in system memory 110. The shortest, and therefore most common, codes are stored in local memory 614, enabling the majority of codeword searches to be performed quickly and efficiently. The next shortest codes

are stored in decoder memory 316 and the longest codes are stored in system memory 110. In this embodiment, the status indicator of each code table entry indicates the memory unit at which to perform the next search if the current search did not result in a codeword match. If the current search did produce a codeword match, the status indicator indicates that condition and the next search will be performed in local memory unit 614. The first search for a data stream, and each search following a codeword match are performed in the local memory unit 614.

In the case of block decoding, the VLD 306 will continue decoding the bitstream as long as there is space available in the block buffer 406. In order to simplify the design, in an illustrative embodiment of the present invention, the VLD 306 checks the buffer availability before starting to decode a block. When the VLD 306 is finished decoding a block, the VLD 306 transfers the data to the block buffer 406. This processing continues until a block count is reached. In an illustrative embodiment, a double buffer scheme is used in order to support high definition (HD) performance.

Referring again to Figure 3, picture-level processing, from the sequence level down to the macroblock level, including the sequence headers, picture headers, time stamps, and buffer management, are performed directly and sequentially by the core processor 302. The VLD 306 assists the core processor when a bit-field in a header is to be decoded. In an illustrative embodiment picture level processing does not overlap with slice level (macroblock) decoding. In an alternative embodiment, some slice level or macroblock decoding processes may be performed concurrently with the picture level processing.

The macroblock level decoding is the main video decoding process. In an illustrative embodiment of the present invention it occurs within a direct execution loop. In such an embodiment, hardware blocks VLD 306, IQ 308, IDCT module 309, pixel filter 310, motion compensation module 312 (and possibly loop filter 313) are all involved in the decoding loop. The core processor 302 controls the loop by polling the status of each of the hardware blocks involved.

In an illustrative embodiment of the present invention, the actions of the various hardware blocks are arranged in an execution pipeline. The pipeline scheme

aims to achieve maximum utilization of the core processor 302. FIG. 9 is a chart representing a decoding pipeline according to an illustrative embodiment of the present invention. The number of pipeline stages may vary depending on the target applications. Due to the selection of hardware elements that comprise the pipeline, the pipeline architecture of the present invention can accommodate substantially any existing or future compression algorithms that fall into the general class of DCT-based, variable-length coded, block-motion compensated algorithms.

The rows of FIG. 9 represent the decoding functions performed as part of the pipeline according to an exemplary embodiment. Variable length decoding 900 is performed by VLD 306. Run length/inverse scan/IQ/mismatch 902 are functions performed by IQ module 308. IDCT operations 904 are performed by IDCT module 309. Pixel filter reference fetch 906 and pixel filter reconstruction 908 are performed by pixel filter 310. Motion compensation reconstruction 910 is performed by motion compensation module 312. The columns of FIG. 9 represent the pipeline stages. The designations MB_i , MB_{i+1} , MB_{i+2} , etc. represent the i^{th} macroblock in a data stream, the $i+1^{\text{st}}$ macroblock in the data stream, the $i+2^{\text{nd}}$ macroblock, and so on. The pipeline scheme supports one pipeline stage per module, wherein any hardware module that depends on the result of another module is arranged in an immediately following MB pipeline stage.

At any given stage in the pipeline, while a given function is being performed on a given macroblock, the next macroblock in the data stream is being worked on by the previous function in the pipeline. Thus, at stage x 912 in the pipeline represented in FIG. 9, variable length decoding 900 is performed on MB_i . Exploded view 920 of the variable length decoding function 900 demonstrates how functions are divided between the core processor 302 and the VLD 306 during this stage, according to one embodiment of the present invention. Exploded view 920 shows that during stage x 912, the core processor 302 decodes the macroblock header of MB_i . The VLD 306 assists the core processor 302 in the decoding of macroblock headers. The core processor 302 also reconstructs the motion vectors of MB_i , calculates the address of the pixel filter reference fetch for MB_i , performs pipeline flow control and checks the status of IQ module 308, IDCT module 309, pixel filter 310 and motion compensator

312 during stage x 912. The hardware blocks operate concurrently with the core processor 302 while decoding a series of macroblocks. The core processor 302 controls the pipeline, initiates the decoding of each macroblock, and controls the operation of each of the hardware accelerators. The core processor firmware checks the status of each of the hardware blocks to determine completion of previously assigned tasks and checks the buffer availability before advancing the pipeline. Each block will then process the corresponding next macroblock. The VLD 306 also decodes the macroblock coefficients of MB_i during stage x . Block coefficient VLC decoding is not started until the core processor 302 decodes the whole macroblock header. Note that the functions listed in exploded view 920 are performed during each stage of the pipeline of FIG. 9, even though, for simplicity's sake, they are only exploded out with respect to stage x 912.

At the next stage $x+1$ 914, the inverse quantizer 308 works on MB_i (function 902) while variable length decoding 900 is performed on the next macroblock, MB_{i+1} . In stage $x+1$ 914, the data that the inverse quantizer 308 works on are the quantized DCT coefficients of MB_i extracted from the data stream by the VLD 306 during stage x 912. In an exemplary embodiment of the present invention, also during stage $x+1$ 914, the pixel filter reference data is fetched for MB_i (function 906) using the pixel filter reference fetch address calculated by the core processor 302 during stage x 912.

Then, at stage $x+2$ 916, the IDCT module 309 performs IDCT operations 904 on the MB_i DCT coefficients that were output by the inverse quantizer 308 during stage $x+1$. Also during stage $x+2$, the pixel filter 310 performs pixel filtering 908 for MB_i using the pixel filter reference data fetched in stage $x+1$ 914 and the motion vectors reconstructed by the core processor 302 in stage x 912. Additionally at stage $x+2$ 916, the inverse quantizer 308 works on MB_{i+1} (function 902), the pixel filter reference data is fetched for MB_{i+1} (function 906), and variable length decoding 900 is performed on MB_{i+2} .

At stage $x+3$ 918, the motion compensation module 312 performs motion compensation reconstruction 910 on MB_i using decoded difference pixel information produced by the IDCT module 309 (function 904) and pixel prediction data produced by the pixel filter 310 (function 908) in stage $x+2$ 916. Also during stage $x+3$ 918,

the IDCT module 309 performs IDCT operations 904 on MB_{i+1} , the pixel filter 310 performs pixel filtering 908 for MB_{i+1} , the inverse quantizer 308 works on MB_{i+2} (function 902), the pixel filter reference data is fetched for MB_{i+2} (function 906), and variable length decoding 900 is performed on MB_{i+3} . While the pipeline of FIG. 9 shows just four pipeline stages, in an illustrative embodiment of the present invention, the pipeline includes as many stages as is needed to decode a complete incoming data stream with adequate performance.

In an alternative embodiment of the present invention, the functions of two or more hardware modules are combined into one pipeline stage, and the macroblock data is processed by all the modules in that stage sequentially. For example, in an exemplary embodiment, IDCT operations for a given macroblock are performed during the same pipeline stage as IQ operations. In this embodiment, the IDCT module 309 waits idle until the inverse quantizer 308 finishes, and the inverse quantizer 308 becomes idle when the IDCT operations start. This embodiment will have a longer processing time for the "packed" pipeline stage, assuming the same performance of each individual function. Therefore, in an illustrative embodiment of the present invention, the packed pipeline stage is used only in non-demanding decoding tasks such as SD (standard definition) or SIF (standard interchange format) size decoding applications. In a further illustrative embodiment using packed stages, different operations such as IQ and IDCT functions are performed by a configurable common block hardware at different times. The benefits of the packed stage embodiment include fewer pipeline stages, fewer buffers and possibly simpler control for the pipeline.

The above-described macroblock-level pipeline advances stage-by-stage. Conceptually, the pipeline advances after all the tasks in the current stage are completed. The time elapsed in one macroblock pipeline stage will be referred to herein as the macroblock (MB) time. In the general case of decoding, the MB time is not a constant and varies from stage to stage. It depends on the encoded bitstream characteristics and possibly other factors, and is determined by the bottleneck module, which is the one that finishes last in that stage. Any module, including the core processor 302 itself, could be the bottleneck from stage to stage, and it is not pre-

determined at the beginning of each stage. In an illustrative embodiment of the present invention, the bottleneck time is reduced by means of firmware control, improving the throughput and directly contributing to performance enhancement.

However, for a given encoding/decoding algorithm, each module, including the core processor 302, has a defined and predetermined task or group of tasks. The maximum number of clock cycles needed for each module to decode decoding of a specific, e.g. worst case, stream can be predetermined. The macroblock time for each module is substantially constant for streams with a given set of characteristics. Therefore, in an illustrative embodiment of the present invention, the hardware acceleration pipeline is optimized by hardware balancing each module in the pipeline according to the compression format of the data stream.

The main video decoding operations occur within a direct execution loop with polling of the accelerator functions. The coprocessor/accelerators operate concurrently with the core processor while decoding a series of macroblocks. The core processor 302 controls the pipeline, initiates the decoding of each macroblock, and controls the operation of each of the accelerators. Upon completion of each macroblock processing stage in the core processor, firmware checks the status of each of the accelerators to determine completion of previously assigned tasks. In the event that the firmware gets to this point before an accelerator module has completed its required tasks, the firmware polls for completion. This is appropriate, since the pipeline cannot proceed efficiently until all of the pipeline elements have completed the current stage, and an interrupt driven scheme would be less efficient for this purpose.

Each hardware module 306, 308, 309, 310, 312, 313, 315 is independently controllable by the core processor 302. The core processor 302 drives a hardware module by issuing a certain start command after checking the module's status. In one embodiment, the core processor 302 issues the start command by setting up a register in the hardware module.

FIG. 10 is a flowchart representing a macroblock decoding loop according to an illustrative embodiment of the present invention. FIG. 10 depicts the decoding of one video picture. In an illustrative embodiment of the present invention, the loop of

slice/macroblock level decoding pipeline control is fully synchronous. At step 1000, the core processor 302 retrieves a macroblock to be decoded from system memory 110. At step 1010, the core processor starts all the hardware modules except the VLD 306. At step 1020, the core processor 302 decodes the macroblock header with the help of the VLD 306. At step 1030, when the macroblock header is decoded, the core processor 302 starts the VLD 306 for block coefficient decoding. At step 1040, the core processor 302 calculates motion vectors and memory addresses, such as the pixel filter reference fetch address, controls buffer rotation and performs other housekeeping tasks. At decision box 1050, if the picture is decoded, the process is complete. If the picture is not decoded, the core processor 302 retrieves the next macroblock, and the process continues as shown by step 1000. In an illustrative embodiment of the present invention, when the current picture has been decoded, the incoming macroblock data of the next picture in the video sequence is decoded according to the process of FIG. 10.

In general, the core processor 302 interprets the bits decoded (with the help of the VLD 306) for the appropriate headers and sets up and coordinates the actions of the hardware modules. More specifically, all header information, from the sequence level down to the macroblock level, is requested by the core processor 302. The core processor 302 also controls and coordinates the actions of each hardware module. The core processor configures the hardware modules to operate in accordance with the encoding/decoding format of the data stream being decoded by providing operating parameters to the hardware modules. The parameters include but are not limited to (using MPEG-2 as an example) the cbp pattern used by the VLD 306 to decode the macroblock coefficients, the quantization scale used by the IQ module 308 to perform inverse quantization, motion vectors used by the pixel filter 309 and motion compensation module 310 to reconstruct the macroblocks, and the working buffer address(es) in decoder memory 316.

Each hardware module 306, 308, 309, 310, 312, 313, 315 performs the specific processing as instructed by the core processor 302 and sets up its status properly in a status register as the task is being executed and when it is done. Each of the modules has or shares a status register that is polled by the core processor to

determine the module's status. Each hardware module is assigned a set of macroblock buffers in decoder memory 316 for processing purposes. Each hardware module signals the busy/available status of the working buffer(s) associated with it so that the core processor 302 can properly coordinate the processing pipeline.

5 In an exemplary embodiment of the present invention, the hardware accelerator modules 306, 308, 309, 319, 312, 313, 315 generally do not communicate with each other directly. The accelerators work on assigned areas of decoder memory 316 and produce results that are written back to decoder memory 316, in some cases to the same area of decoder memory 316 as the input to the accelerator. In one
10 embodiment of the present invention, when the incoming bitstream is of a format that includes a relatively large amount of data, or of a relatively complex encoding/decoding format, the accelerators in some cases may bypass the decoder memory 316 and pass data between themselves directly.

Software codes from other sources, such as proprietary codes, are ported to the
15 decoding system 300 by analyzing the code to isolate those functions that are amenable to acceleration, such as variable-length decoding, run-length coding, inverse scanning, inverse quantization, transform, pixel filter, motion compensation, de-blocking filter, and display format conversion, and replacing those functions with equivalent functions that use the hardware accelerators in the decoding system 300.
20 All other video decoding software is compiled to run directly on the core processor.

FIG. 11 is a functional block diagram of a digital video decoding system 1100 according to an illustrative embodiment of the present invention. Video decoding system 1100 is similar to the video decoding system 300 shown in FIG. 3, but includes two variable-length decoders, VLD₀ 1110 and VLD₁ 1120. The other
25 elements of FIG. 11 are equivalent to the elements shown and described with respect to FIG. 3. In an illustrative embodiment, both of the variable-length decoders 1110 and 1120 are programmable to decode bitstreams of a plurality of compression/decompression standards. In this case, each of the variable-length decoders 1110 and 1120 has a code RAM to hold VLC tables for various video
30 coding formats. In an alternative embodiment, one of the variable-length decoders 1110 and 1120 is programmable to operate according to a plurality of

compression/decompression standards, and the other variable-length decoder is
hardwired to efficiently perform decoding according to one or more particular
compression standards, such as MPEG-2 HD. In another embodiment, both of the
variable-length decoders 1110 and 1120 are hardwired to efficiently perform decoding
5 according to one or more particular compression standards. In still another
embodiment, one or both of the two VLDs 1110 and 1120 is hardwired to decode
bitstreams according to one or more particular standards and can also be optionally set
as a programmable VLD engine, with a code RAM to hold VLC tables for other video
coding formats. In the embodiments wherein a variable-length decoder is hard-wired,
10 the VLD includes a hard-coded coefficient decoder and a hard-coded code look-up
table. The two VLD engines 1110 and 1120 are controlled independently by the core
processor 302, and either one or both of them will be employed at any given time,
depending on the application.

In an exemplary embodiment of the present invention, the two variable-length
15 decoders 1100 and 1110 are employed concurrently to decode the same bitstream. In
one embodiment, the two variable-length decoders are used in an alternating fashion
to decode incoming macroblocks. That is VLD₀ 1110 is used to decode a first
macroblock, VLD₁ 1120 is used to decode a second macroblock, VLD₀ 1100 is used
to decode the third macroblock, and so on. In an illustrative embodiment, two rows of
20 a video frame are decoded concurrently, with one row being decoded by one VLD,
and the other row being decoded by the other VLD. FIG. 12 is a chart representing a
decoding pipeline according to an illustrative embodiment of the present invention.
The rows of FIG. 12 represent the functions performed by the core decoder processor
302, VLD₀ 1110 and VLD₁ 1120 as part of the pipeline, according to an exemplary
25 embodiment. Row 1200 shows the functions performed by the core processor 302,
row 1202 shows the functions performed by VLD₀ 1110, and row 1204 shows the
functions performed by VLD₁ 1120. The columns of FIG. 12 represent the pipeline
stages.

As can be seen in FIG. 12, the variable-length decoding of each macroblock is
30 performed by one of the two VLDs 1110 and 1120 over two decoding stages. For a
given macroblock, in a first stage, the assigned VLD assists the core processor 302 in

decoding the macroblock header. In the next stage, the same VLD performs macroblock coefficient decoding for the same macroblock, while the other VLD assists the core processor 302 in decoding the macroblock header of a macroblock in a different row. In the example of FIG. 12, at stage x 1206, the decoder processor 302 performs macroblock header decoding on the first macroblock of row₀ (row₀, column₀). Simultaneously, VLD₀ 1110 assists the core processor 302 in the decoding of the macroblock header of the same macroblock (row₀, column₀) data. At the next stage $x+1$ 1208, the decoder processor 302 performs macroblock header decoding on the first macroblock of row₁ (row₁, column₀), and VLD₁ 1120 assists the core processor 302 in the decoding of the macroblock header of the same macroblock (row₁, column₀) data. Also during stage $x+1$, VLD₀ 1110 performs macroblock coefficient decoding for the first macroblock of row₀ (row₀, column₀), whose header was decoded by VLD₀ in stage x . At stage $x+2$ 1208, the decoder processor 302 performs macroblock header decoding on the second macroblock of row₀ (row₀, column₁), and VLD₀ 1110 assists the core processor 302 in the decoding of the macroblock header of the same macroblock (row₀, column₁) data. Also, during stage $x+2$, VLD₁ 1120 performs macroblock coefficient decoding for the first macroblock of row 1 (row₁, column₀), whose header was decoded by VLD₁ in stage $x+1$. Decoding continues in this manner.

In the decoding process depicted in FIG. 12, after the macroblock header is decoded for a given macroblock, coefficient decoding for that macroblock is not initiated until the next stage. In an alternative embodiment of the present invention, the variable length decoder that is working on a given macroblock does not wait for the next stage after assisting the core processor 302 in decoding the macroblock header. Rather, when the decoding of the macroblock header is complete, the variable-length decoder begins decoding the macroblock coefficients for that macroblock, regardless of whether or not the next stage is ready to begin. This process will be referred to as cycle stealing.

FIG. 13 is a chart representing a dual-row decoding pipeline employing cycle stealing according to an illustrative embodiment of the present invention. The rows of FIG. 13 represent the decoding functions performed as part of the pipeline,

according to an exemplary embodiment of the present invention. The functions include core processor operations 1300, variable-length decoding performed by VLD₀ 1302, variable-length decoding performed by VLD₁ 1304, inverse quantizer operations 1306, IDCT operations 1308, pixel filter reference fetch 1310, pixel filter reconstruction 1312, motion compensation 1314 and DMA operations 1316. The columns of FIG. 13 represent the pipeline stages. The designation (i, j) denotes the macroblock coordinates, i.e., the jth MB in the ith row.

As shown in FIG. 13, in stage 1, the core processor 302 and VLD₀ 1110 work on MB_{0,0} (MB₀ in row₀). Note that, first, the core processor 302 performs macroblock header decoding with the assistance of VLD₀ 1110. When the macroblock header is decoded, the core processor 302 continues performing other tasks, while VLD₀ 1110 begins decoding the block coefficients of MB_{0,0}. When the core processor 302 completes the tasks that it is performing with respect to MB_{0,0}, the core processor 302 initiates stage 2, regardless of whether VLD₀ 1110 has finished decoding the block coefficients of MB_{0,0}. In an alternative embodiment of the present invention, after assisting the core processor 302 with decoding the macroblock header, VLD₀ 1110 waits until stage 2 to begin decoding the block coefficients of MB_{0,0}, as depicted in FIG. 12.

In stage 2, the core processor 302 and VLD₁ 1120 work on MB_{1,0} (MB₀ in row₁). First the core processor 302 performs macroblock header decoding on MB_{1,0} with the assistance of VLD₁ 1120. When the macroblock header is decoded, the core processor 302 continues performing other tasks while VLD₁ 1120 begins decoding the block coefficients of MB_{1,0}. Also in stage 2, if VLD₀ 1110 did not finish decoding the block coefficients of MB_{0,0} in stage 1, it (VLD₀ 1110) continues to do so in stage 2. In the alternative embodiment mentioned above with respect to stage 1, VLD₀ 1110 waits until stage 2 to begin decoding the block coefficients of MB_{0,0}. When the core processor 302 completes the tasks that it is performing with respect to MB_{1,0}, the core processor 302 polls VLD₀ 1110 to see if it is done decoding the block coefficients of MB_{0,0}. If VLD₀ 1110 is done with MB_{0,0}, the core processor 302 initiates stage 3, regardless of whether VLD₁ 1120 has finished decoding the block coefficients of MB_{1,0}. If VLD₀ 1110 is not yet finished decoding the block coefficients of MB_{0,0}, the

core processor 302 waits until VLD₀ 1110 is finished with MB_{0,0} and initiates stage 3 at that time, again, regardless of whether VLD₁ 1120 has finished decoding the block coefficients of MB_{1,0}.

5 In stage 3, the core processor 302 and VLD₀ 1110 work on MB_{0,1} (MB₁ in row₀) as described above with respect to stages 1 and 2. Also in stage 3, IQ module 308 operates on MB_{0,0}, performing run-level code decoding, inverse scanning, inverse quantization and mismatch control. The data that the inverse quantizer 308 works on are the quantized DCT coefficients of MB_{0,0}, extracted from the data stream by VLD₀ 1110 during stage 2. Additionally in stage 3, VLD₁ 1120 continues decoding the
10 block coefficients of MB_{1,0}, if the decoding was not completed in stage 2. When the core processor 302 completes the tasks that it is performing with respect to MB_{0,1}, the core processor 302 polls VLD₁ to see if it is done decoding the block coefficients of MB_{1,0}. The core processor 302 also polls IQ module 308 to see if it is done operating on MB_{0,1}. If VLD₁ 1120 is done with MB_{0,0}, and IQ module 308 is done with MB_{0,1},
15 the core processor 302 initiates stage 4, regardless of whether VLD₀ 1110 has finished decoding the block coefficients of MB_{0,1}. If either VLD₁ 1120 or IQ module 308 is not yet finished, the core processor 302 waits until VLD₁ 1120 and IQ module 308 are both finished and initiates stage 4 at that time. In an exemplary embodiment of the present invention, also during stage 3, the pixel filter reference data is fetched for
20 MB_{0,0} (function 910), using the pixel filter reference fetch address calculated by the core processor 302 during stage 1. In this case, the core processor 302 also polls the pixel filter 310 for completion prior to initiating stage 4.

25 In stage 4, the core processor 302 works on MB_{1,1} (MB₁ in row₁), variable-length decoding is initiated on MB_{1,1} by VLD₁ 1120, IQ module 308 operates on MB_{1,0}, and the pixel filter reference data is fetched for MB_{1,0} (function 910). Also in stage 4, IDCT module 309 performs the inverse transform on the MB_{0,0} coefficients produced by the IQ module 308 in stage 3, and the pixel filter 310 performs pixel filtering 912 for MB_{0,0}, using the pixel filter reference data fetched in stage 3 and the motion vectors reconstructed by the core processor 302 in stage 1. Additionally in
30 stage 4, VLD₀ 1110 continues decoding the block coefficients of MB_{0,1} if the decoding was not completed in stage 3. When the core processor 302 completes its

tasks with respect to $MB_{1,1}$, the core processor 302 polls VLD_0 1110, IQ module 308, IDCT module 309 and pixel filter 310 to see if they have completed their present tasks. If the polled modules have completed their tasks, the core processor 302 initiates stage 5. If any of the polled modules is not yet finished, the core processor
5 waits until they are all finished and initiates stage 5 at that time.

In stage 5, the core processor 302 works on $MB_{0,2}$ (MB_2 in row₀), variable-length decoding is initiated on $MB_{0,2}$ by VLD_0 1110, IQ module 308 operates on $MB_{0,1}$, IDCT module 309 operates on the $MB_{1,0}$ coefficients, the pixel filter reference data is fetched for $MB_{0,1}$ (function 910), and the pixel filter 310 performs pixel
10 filtering 912 for $MB_{1,0}$. Also in stage 5, the motion compensation module 312 performs motion compensation reconstruction 914 on $MB_{0,0}$, using decoded difference pixel information produced by the IDCT module 309 (function 908) and pixel prediction data produced by the pixel filter 310 (function 912) in stage 4 616. Additionally, in stage 5, VLD_1 1120 continues decoding the block coefficients of
15 $MB_{1,1}$ if the decoding was not completed in stage 4. When the core processor 302 completes its tasks with respect to $MB_{0,2}$, the core processor 302 polls VLD_1 1120, IQ module 308, IDCT module 309, pixel filter 310 and motion compensation module 312 to see if they have completed their present tasks. If the polled modules have completed their tasks, the core processor 302 initiates stage 6. If any of the polled
20 modules is not yet finished, the core processor waits until they are all finished and initiates stage 6 at that time.

In stage 6, the core processor 302 works on $MB_{1,2}$ (MB_2 in row₁), variable-length decoding is initiated on $MB_{1,2}$ by VLD_1 1120, IQ module 308 operates on $MB_{1,1}$, IDCT module 309 operates on the $MB_{0,1}$ coefficients, the pixel filter reference data is fetched for $MB_{1,1}$ (function 910), the pixel filter 310 performs pixel filtering
25 912 for $MB_{0,1}$ and the motion compensation module 312 performs motion compensation reconstruction 914 on $MB_{1,0}$. Also in stage 6, the DMA engine 304 places the result of the motion compensation performed with respect to $MB_{0,0}$ in system memory 110. Additionally in stage 5, VLD_0 1110 continues decoding the
30 block coefficients of $MB_{0,2}$ if the decoding was not completed in stage 5. When the core processor 302 completes its tasks with respect to $MB_{1,2}$, the core processor 302

polls VLD₁ 1120, IQ module 308, IDCT module 309, pixel filter 310, motion compensation module 312 and DMA engine 304 to see if they have completed their present tasks. If the polled modules have completed their tasks, the core processor 302 initiates stage 7. If any of the polled modules is not yet finished, the core processor waits until they are all finished and initiates stage 7 at that time.

The decoding pipeline described above with respect to FIG. 13 continues as long as there are further macroblocks in the data stream to decode. The dual-row decoding pipeline demonstrated in FIG. 13 can be implemented in any type of decoding scheme (including, e.g., audio decoding) employing any combination of acceleration modules.

Although a preferred embodiment of the present invention has been described, it should not be construed to limit the scope of the appended claims. For example, the present invention is applicable to any type of data utilizing variable-length code, including any media data, such as audio data and graphics data, in addition to the video data illustratively described herein. Those skilled in the art will understand that various modifications may be made to the described embodiment. Moreover, to those skilled in the various arts, the invention itself herein will suggest solutions to other tasks and adaptations for other applications. It is therefore desired that the present embodiments be considered in all respects as illustrative and not restrictive, reference being made to the appended claims rather than the foregoing description to indicate the scope of the invention.

WHAT IS CLAIMED IS:

1. A video decoding system comprising:

a decoder processor adapted to perform decoding functions on a video

5 data stream;

a first variable-length decoding accelerator coupled to the decoder processor and adapted to perform variable-length decoding operations on variable-length code in the video data stream; and

10 a second variable-length decoding accelerator coupled to the decoder processor and adapted to perform variable-length decoding operations on variable-length code in the video data stream;

wherein the first and second variable-length decoding accelerators are adapted to cooperatively decode variable-length codes in the video data stream.

15 2. The system of claim 1 wherein the first and second variable-length decoding accelerators are adapted to alternately decode variable-length code data elements in the video data stream.

20 3. The system of claim 1 wherein the first and second variable-length decoding accelerators are adapted to decode variable-length code data elements from the video data stream in parallel.

25 4. The system of claim 1 wherein the variable-length code data elements comprise a sequence of macroblock data elements, each macroblock data element representing a macroblock of a video frame.

30 5. The system of claim 4 wherein each macroblock data element comprises a macroblock header and coefficient data, wherein the first and second variable-length decoding accelerators are adapted to alternately decode macroblock data elements in the video data stream.

6. The system of claim 5 wherein the first and second variable-length decoding accelerators are adapted such that the first variable-length decoding accelerator decodes a macroblock header of one macroblock data element while the second variable-length decoding accelerator decodes coefficient data of another macroblock data element, and the second variable-length decoding accelerator decodes a macroblock header of one macroblock data element while the first variable-length decoding accelerator decodes coefficient data of another macroblock data element.

7. The system of claim 5 wherein each of the variable-length decoding accelerators is adapted to decode the macroblock header of each macroblock data element before decoding the coefficient data of the macroblock data element.

8. The system of claim 5 wherein the variable-length decoding accelerators are adapted such that the decoding of successive macroblock data elements in the data stream is initiated during corresponding successive variable time periods.

9. The system of claim 8 wherein the variable-length decoding accelerators are adapted such that the decoding of each macroblock data element in the data stream is completed before the end of the variable time period that follows the variable time period in which decoding of that macroblock data element was initiated.

10. A video decoding system comprising:
a decoder processor adapted to perform decoding functions on a video data stream;

a first variable-length decoding accelerator coupled to the decoder processor and adapted to perform variable-length decoding operations on variable-length code in the video data stream, wherein the first variable-length decoding accelerator is capable of decoding variable-length codes according to any of a plurality of decoding methods; and

a second variable-length decoding accelerator coupled to the decoder processor and adapted to perform variable-length decoding operations on variable-length code in the video data stream.

5 11. The system of claim 10 wherein the first variable-length decoding accelerator comprises a plurality of code tables stored in memory, each code table corresponding to one of a plurality of sets of variable-length codes, wherein each of the code tables matches variable-length codes to their corresponding decoded information.

10 12. The system of claim 11 wherein the first variable-length decoding accelerator further comprises a register that dictates which of the plurality of code tables is to be utilized to decode variable-length code, wherein the register is programmable to dictate the appropriate code table to be employed.

15 13. The system of claim 10 wherein the second variable-length decoding accelerator is hard-wired to decode variable-length code according to a particular decoding method.

20 14. The system of claim 10 wherein the second variable-length decoding accelerator is capable of decoding variable-length code according to any of a plurality of decoding methods.

25 15. A video decoding system comprising:
a decoder processor adapted to perform decoding functions on a video data stream;

a first variable-length decoding accelerator coupled to the decoder processor and adapted to perform variable-length decoding operations on variable-length code in the video data stream;

30 a second variable-length decoding accelerator coupled to the decoder processor and adapted to perform variable-length decoding operations on variable-length code in the video data stream; and

an inverse quantization accelerator coupled to the decoder processor and adapted to perform inverse quantization operations on data from the video data stream.

5 16 The system of claim 15 further comprising an inverse transform accelerator coupled to the decoder processor and adapted to perform inverse transform operations on data from the video data stream.

10 17. The system of claim 16 wherein the system comprises exactly one decoder processor, one inverse quantization accelerator and one inverse transform accelerator.

15 18. A variable-length decoder comprising:
 a plurality of code tables stored in memory, each code table corresponding to one of a plurality of sets of variable length codes, wherein each of the code tables matches variable-length codes to their corresponding decoded information; and

20 a register adapted to hold a value that dictates which of the plurality of code tables is to be utilized to decode variable-length code, wherein the register is programmable to dictate the appropriate code table to be employed.

25 20. The variable-length decoder of claim 19 further comprising an address generator adapted to generate a memory address at which to search one of the plurality of code tables for a codeword match, wherein the address generator uses a starting address of a code table to be used in order to generate the address at which to search for a codeword match, wherein the register is adapted to hold the starting address of the code table to be used, and wherein the value held in the register is provided to the address generator for use in generating the address at which to search for a codeword match.

30 21. A video decoding system comprising:

a decoder processor adapted to perform decoding functions on a video data stream;

a variable-length decoding accelerator coupled to the decoder processor and adapted to perform variable-length decoding operations on variable-length code in the video data stream, wherein the variable-length decoding accelerator is capable of decoding variable-length code according to any of a plurality of decoding methods.

22. The system of claim 21 wherein the variable-length decoding accelerator comprises a plurality of code tables stored in memory, each code table corresponding to one of a plurality of sets of variable length codes, wherein each of the code tables matches variable-length codes to their corresponding decoded information.

23. The system of claim 22 wherein the variable-length decoding accelerator further comprises a register that dictates which of the plurality of code tables is to be utilized to decode variable-length code, wherein the register is programmable to dictate the appropriate code table to be employed.

24. A video decoding system comprising:

a decoder processor adapted to perform decoding functions on a video data stream;

a variable-length decoding accelerator coupled to the decoder processor and adapted to perform variable-length decoding operations on variable-length code in the video data stream, wherein the variable-length decoding accelerator is hard-wired to decode variable-length code according to a particular decoding method and wherein the variable-length decoding accelerator is configurable to be able to decode variable-length codes according to any of a plurality of decoding methods.

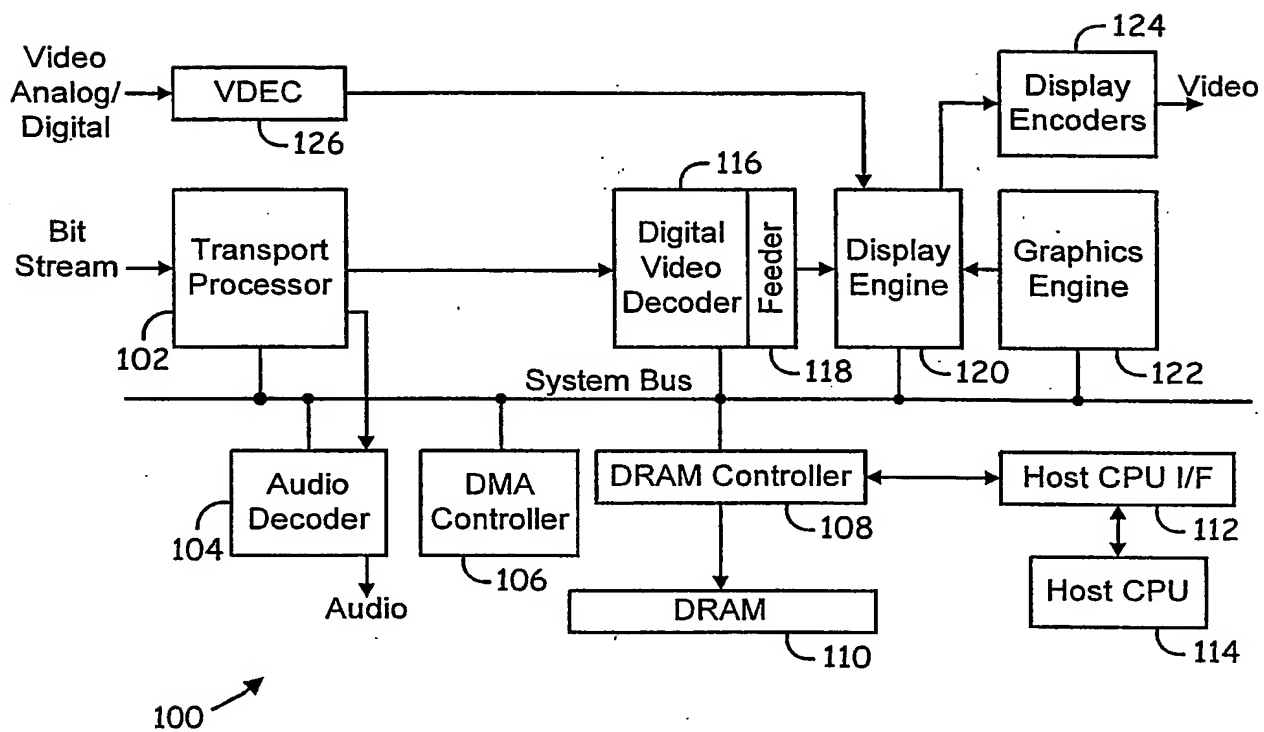


FIG. 1

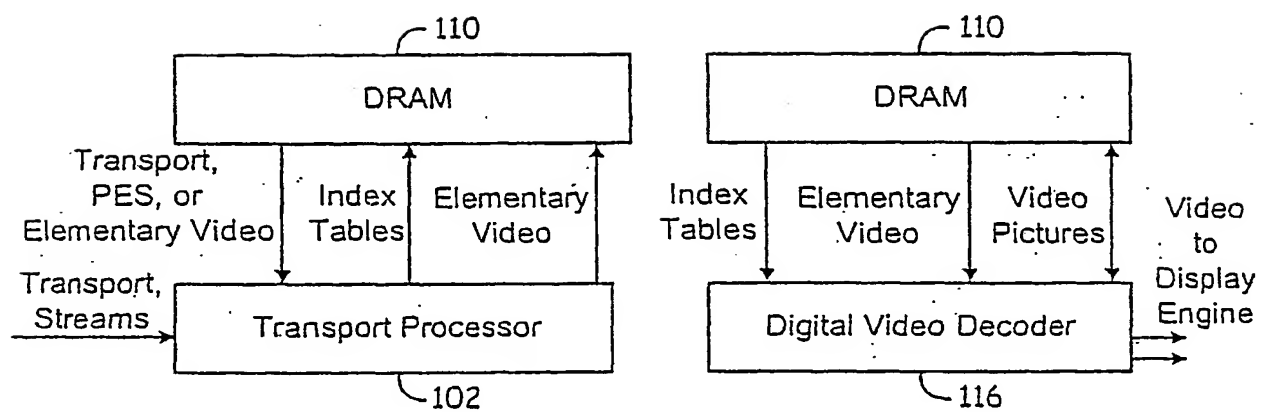


FIG. 2

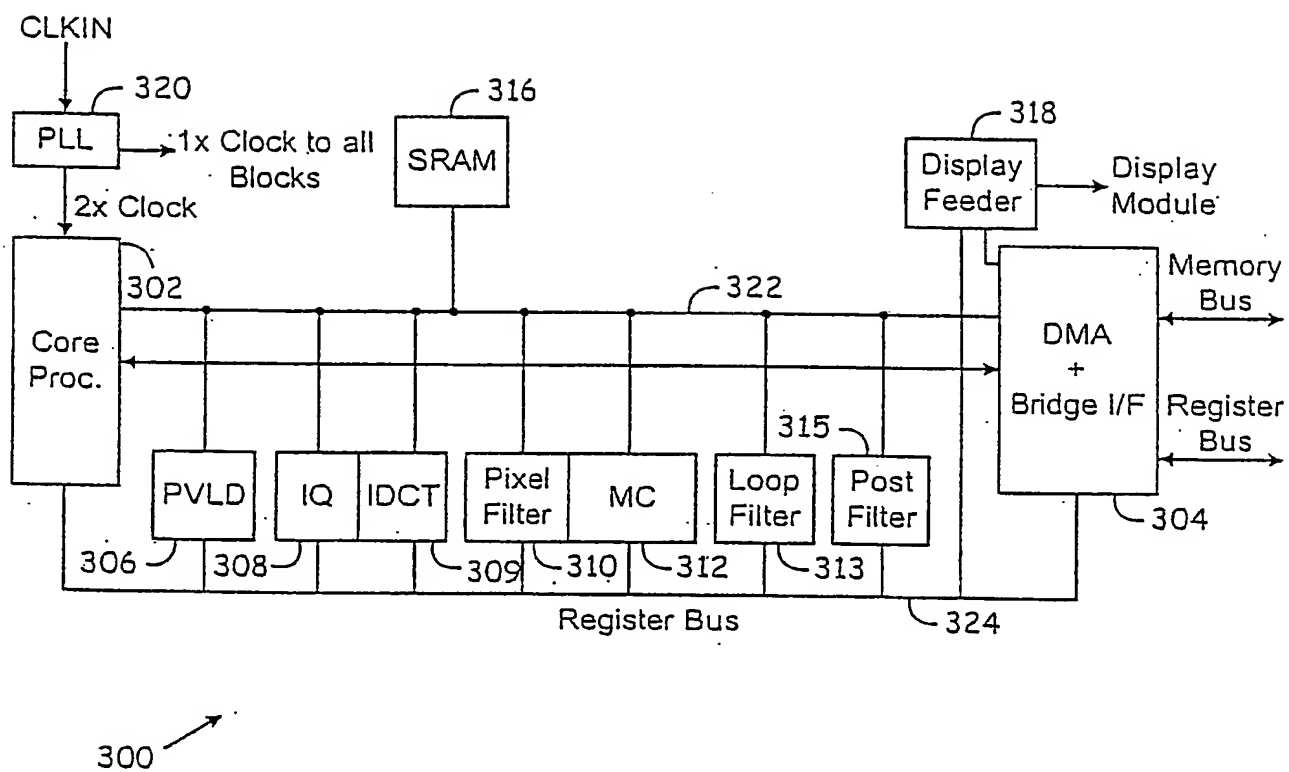


FIG. 3

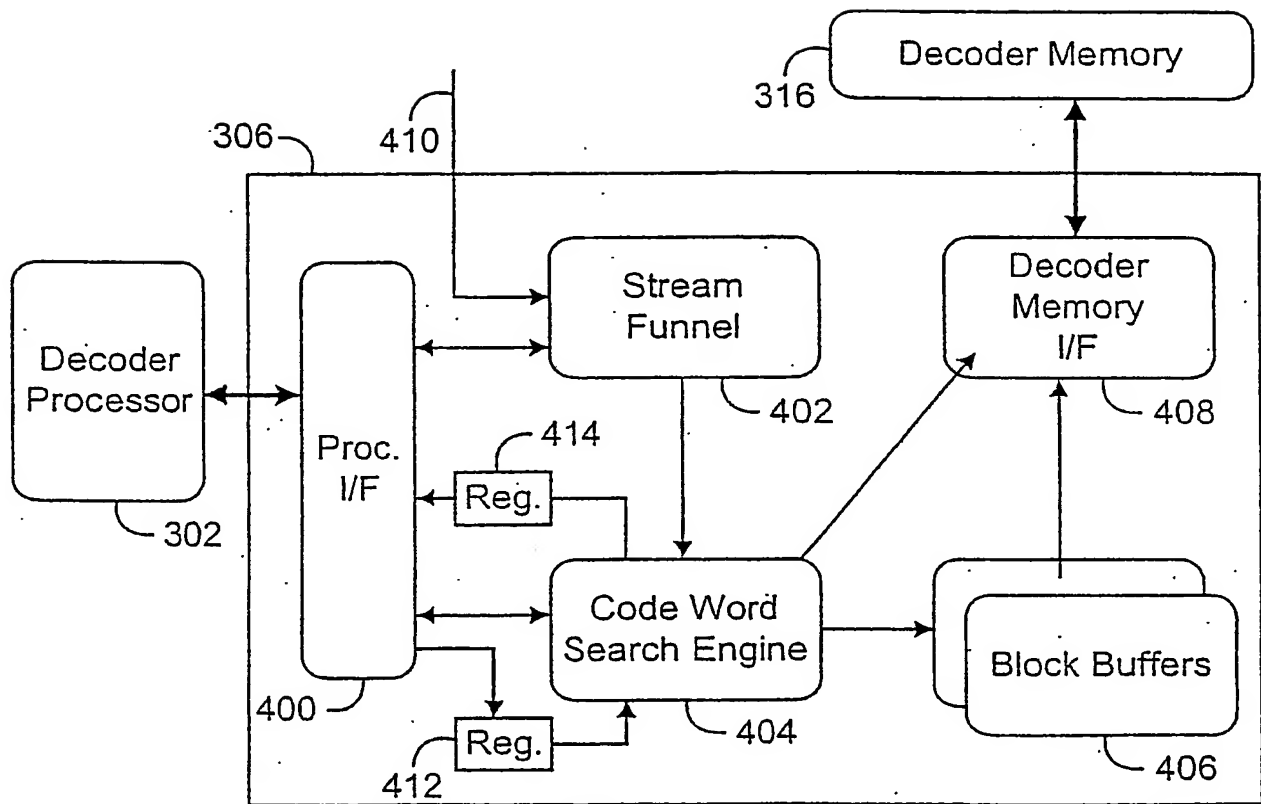


FIG. 4

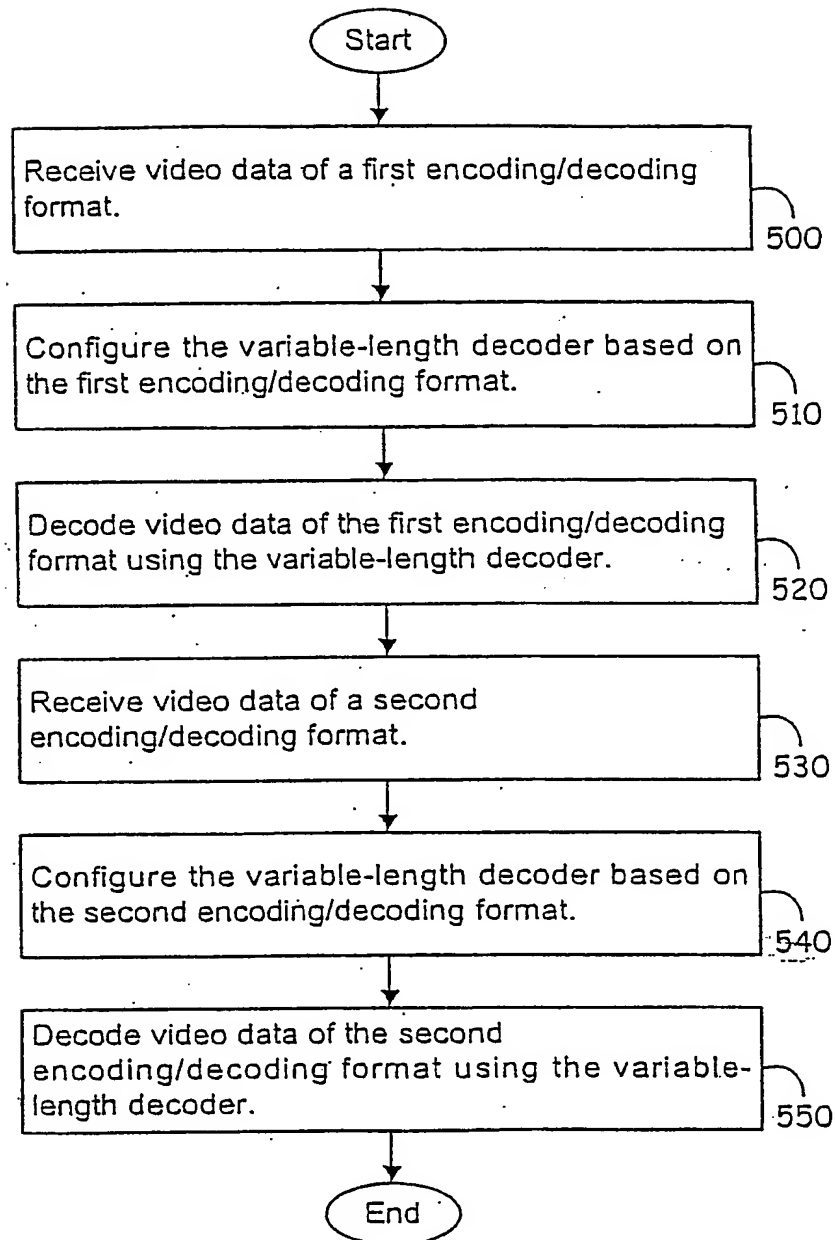


FIG. 5

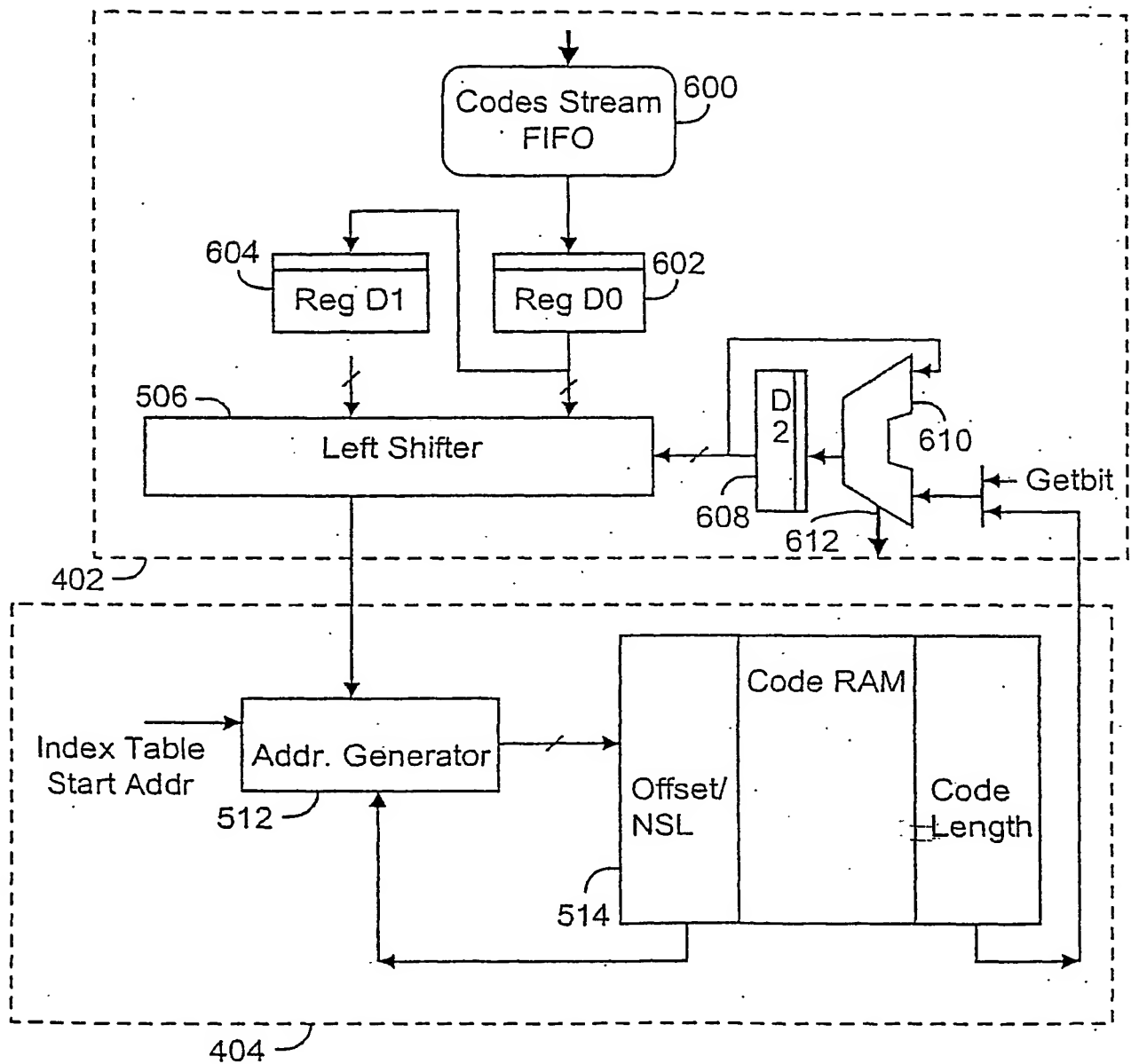


FIG. 6

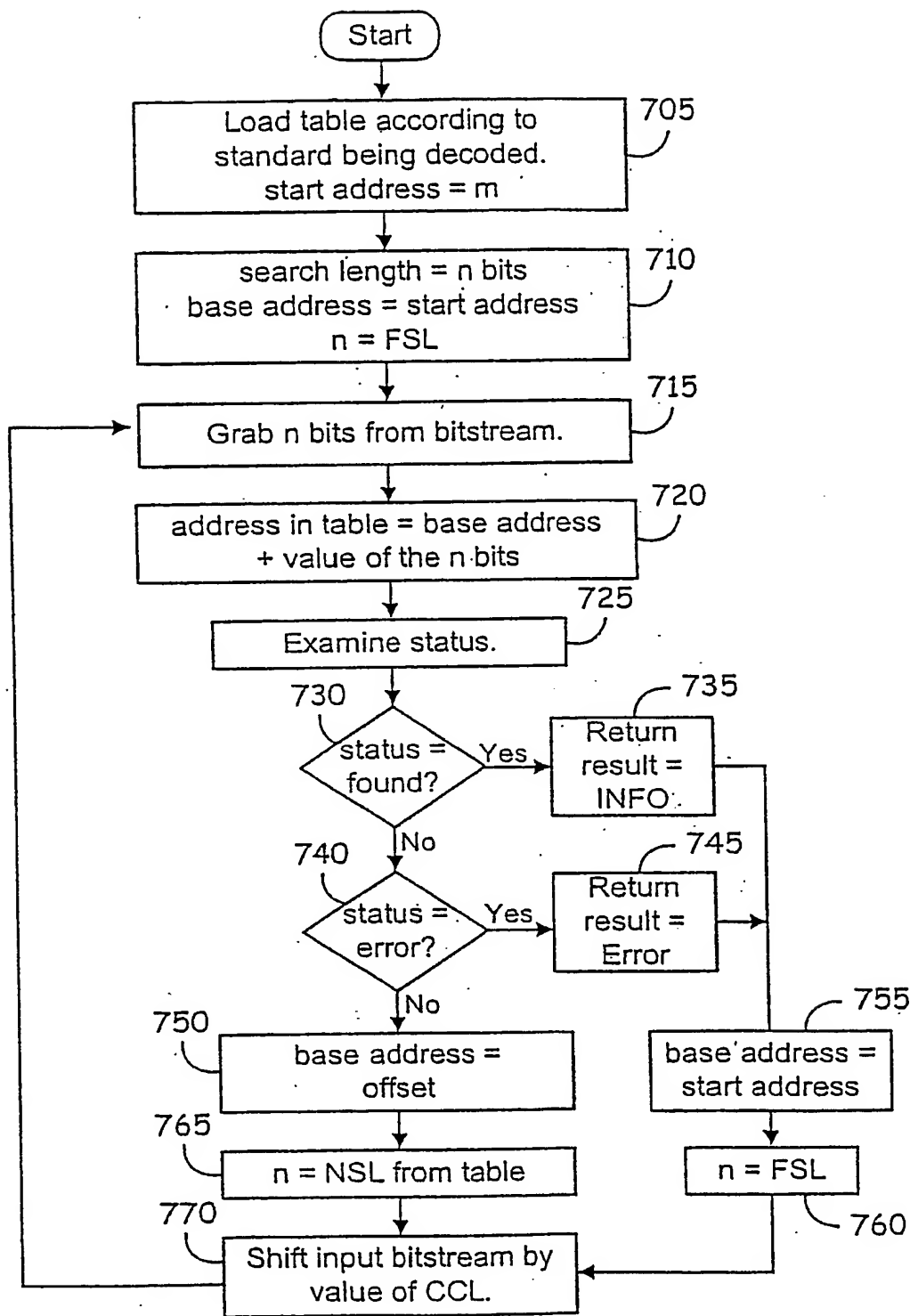


FIG. 7

	800	810	820	830	840	850
Addr #	CCL	NSL	EOB	Status	INFO/OFFSET	
0	3	3	0	0101	8 (offset)	
1	0	0	0	0111 (error)	
2	3	-	0	0100 (found)	B	
3	3	-	0	0100 (found)	C	
4	1	-	0	0100 (found)	A	
5	1	-	0	0100 (found)	A	
6	1	-	0	0100 (found)	A	
7	1	-	0	0100 (found)	A	
....						
12	2	-	0	0100 (found)	X	
13	2	-	0	0100 (found)	X	
14	3	-	0	0100 (found)	Y	
15	3	-	0	0100 (found)	Z	

FIG.8

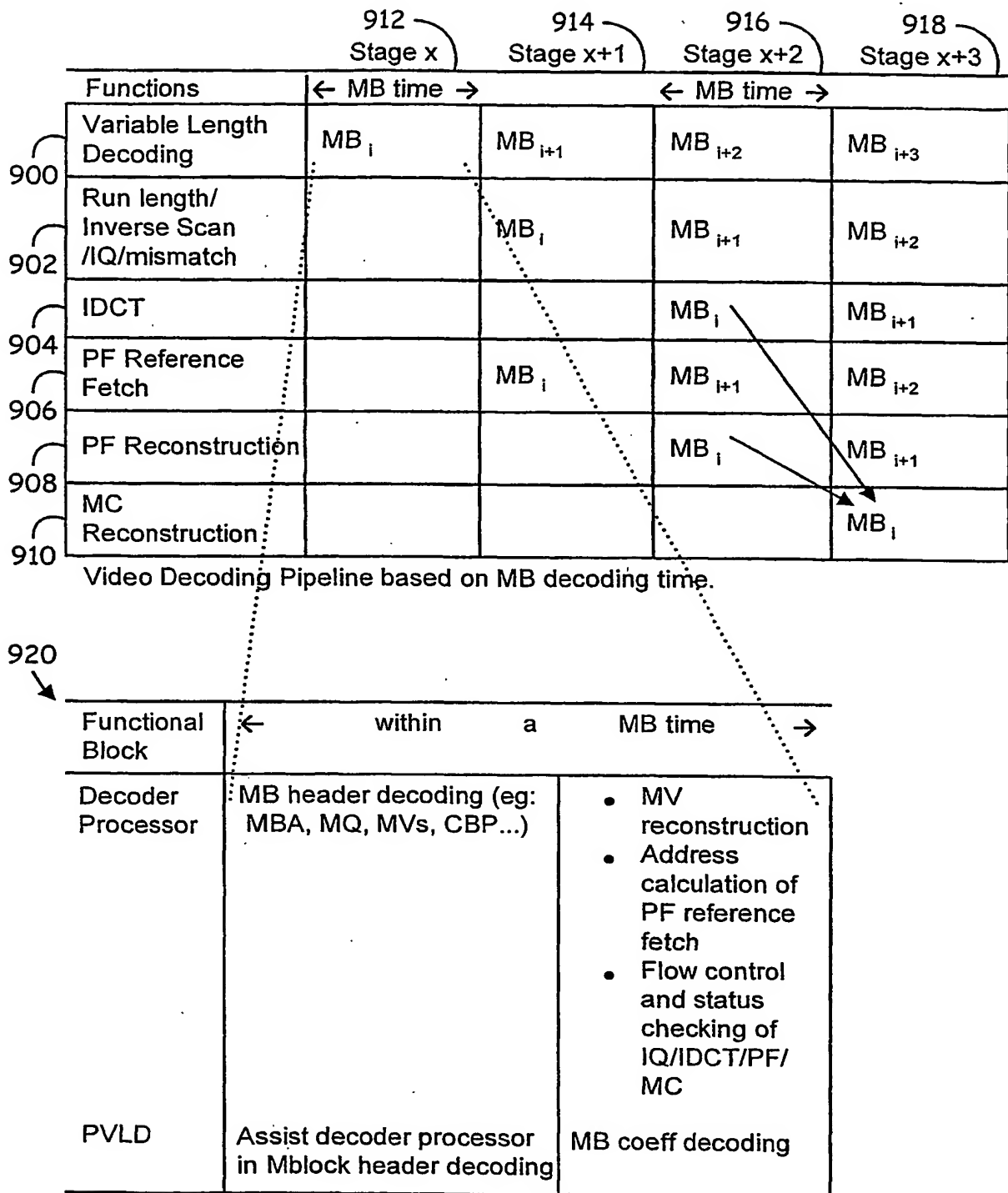


FIG. 9

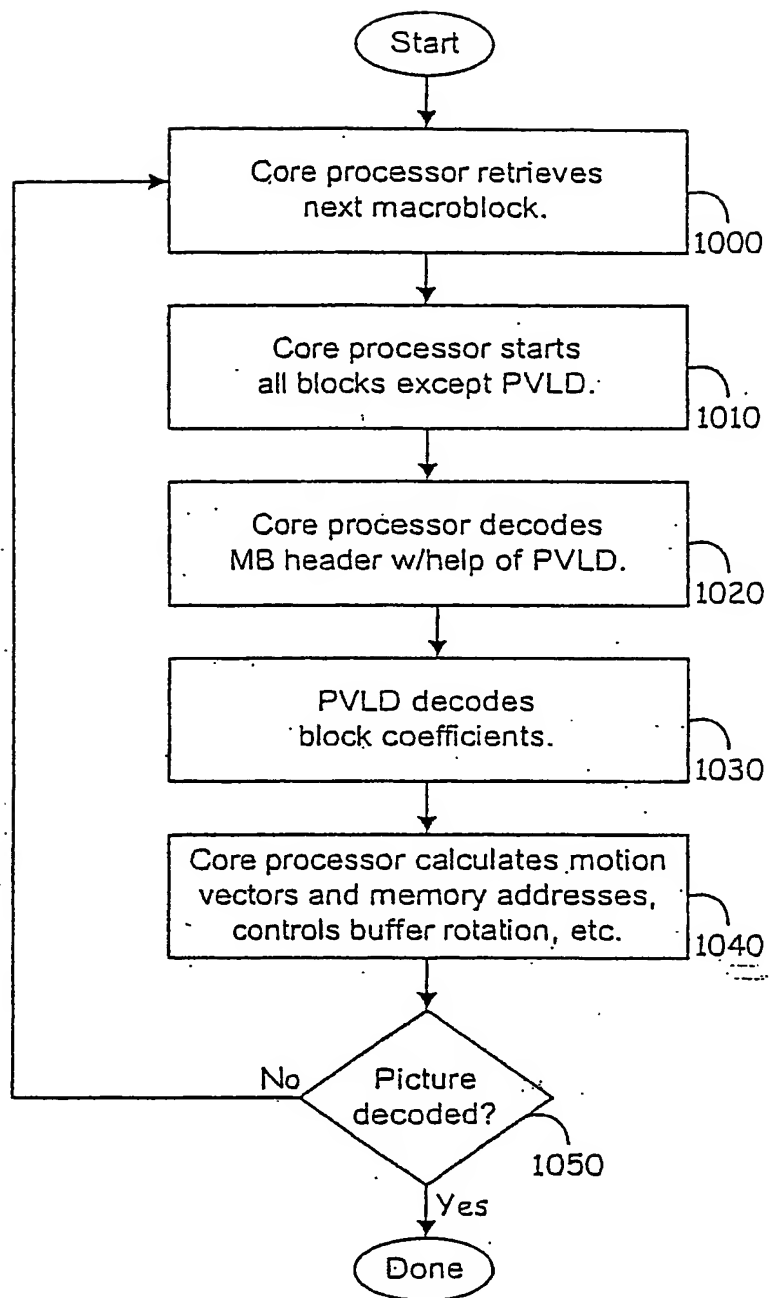


FIG. 10

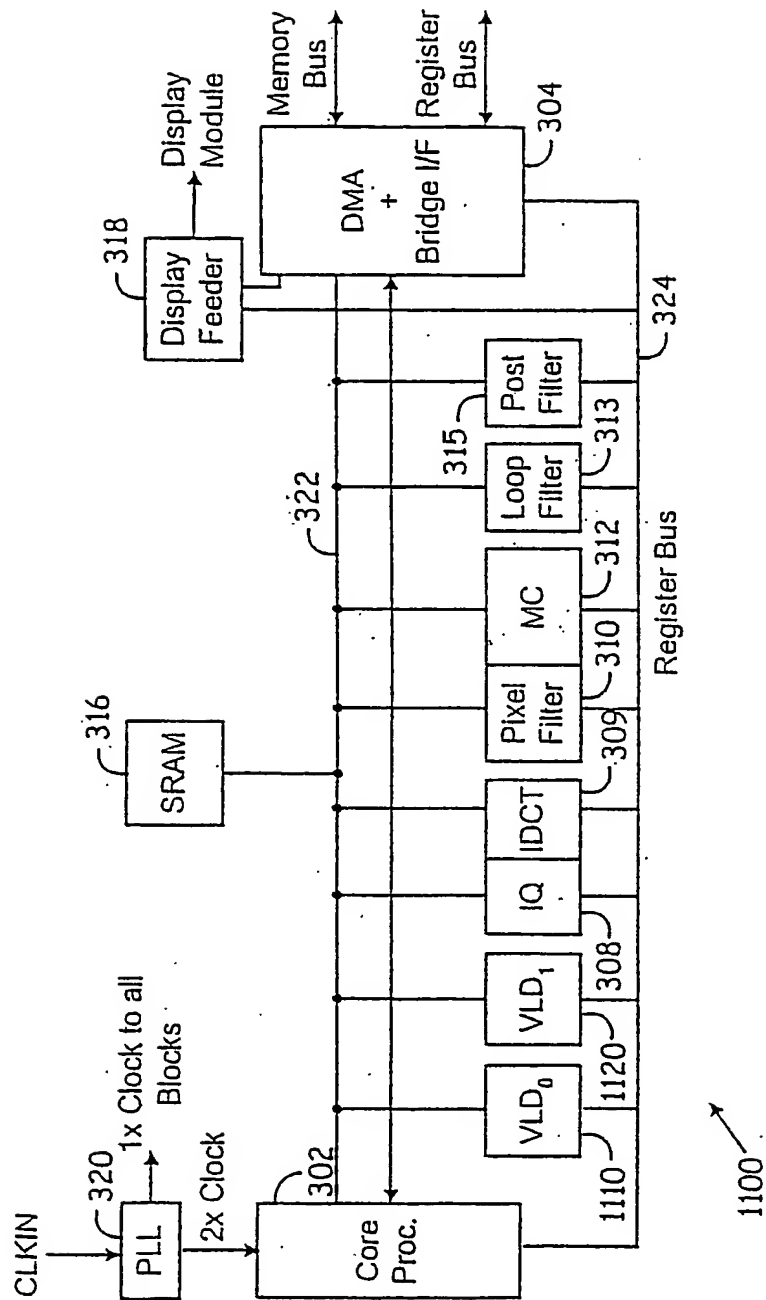


FIG. 11

	Stage x 1206	Stage x+1 1208	Stage x+2 1210
Functional Block	← MB time →	← MB time →	← MB time →
Decoder Processor <u>1200</u>	MB (row ₀ , col ₀) header decoding (eg: MBA, MQ MVs, CBP...)	MB (row ₁ , col ₀) header decoding (eg: MBA, MQ, MVs, CBP...)	MB (row ₀ , col ₁) header decoding (eg: MBA, MQ, MVs, CBP...)
VLD0 <u>1202</u>	Assist decoder processor with MB (row ₀ , col ₀) header decoding	Perform MB (row ₀ , col ₀) coefs decoding	Assist decoder processor with MB (row ₀ , col ₁) header decoding
VLD1 <u>1204</u>		Assist decoder processor with MB (row ₁ , col ₀) header decoding	Perform MB (row ₁ , col ₀) coefs decoding

FIG.12

		Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7	Stage 8
1300	Core Proc.	(0,0)	(1,0)	(0,1)	(1,1)	(0,2)	(1,2)	(0,3)	(1,3)
1302	VLD0	(0,0)	(0,0)	(0,1)	(0,1)	(0,2)	(0,2)	(0,3)	(0,3)
1304	VLD1		(1,0)	(1,0)	(1,1)	(1,1)	(1,2)	(1,2)	(1,3)
1306	IQ			(0,0)	(1,0)	(0,1)	(1,1)	(0,2)	(1,2)
1308	IDCT				(0,0)	(1,0)	(0,1)	(0,2)	(0,2)
1310	PF_REF			(0,0)	(1,0)	(0,1)	(1,1)	(0,1)	(1,2)
1312	PF_REC				(0,0)	(1,0)	(0,1)	(1,1)	(0,2)
1314	MC					(0,0)	(1,0)	(0,1)	(1,1)
1316	DMA						(0,0)	(1,0)	(0,1)

FIG.13

THIS PAGE BLANK (USPTO)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
16 October 2003 (16.10.2003)

PCT

(10) International Publication Number
WO 03/085494 A3

(51) International Patent Classification⁷: **H03M 7/40**

(21) International Application Number: PCT/US03/09896

(22) International Filing Date: 1 April 2003 (01.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/369,144 1 April 2002 (01.04.2002) US

(71) Applicant: **BROADCOM CORPORATION** [US/US];
16215 Alton Parkway, Irvine, CA 92618 (US).

(72) Inventors: **HSIUN, Vivian**; 5580 Avenida El Cid, Yorba Linda, CA 92887 (US). **MACINNIS, Alexander, G.**; 1023 Los Altos Avenue, Los Altos, CA 94022-1017 (US). **XIE, Xiaodong**; 39120 Argonaut Way #472, Fremont, CA 94538 (US).

(74) Agent: **WIBERG, John, A.**; McAndrews, Held & Malloy, Ltd., 500 West Madison, Suite 3400, Chicago, IL 60661 (US).

(84) Designated States (*regional*): European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR).

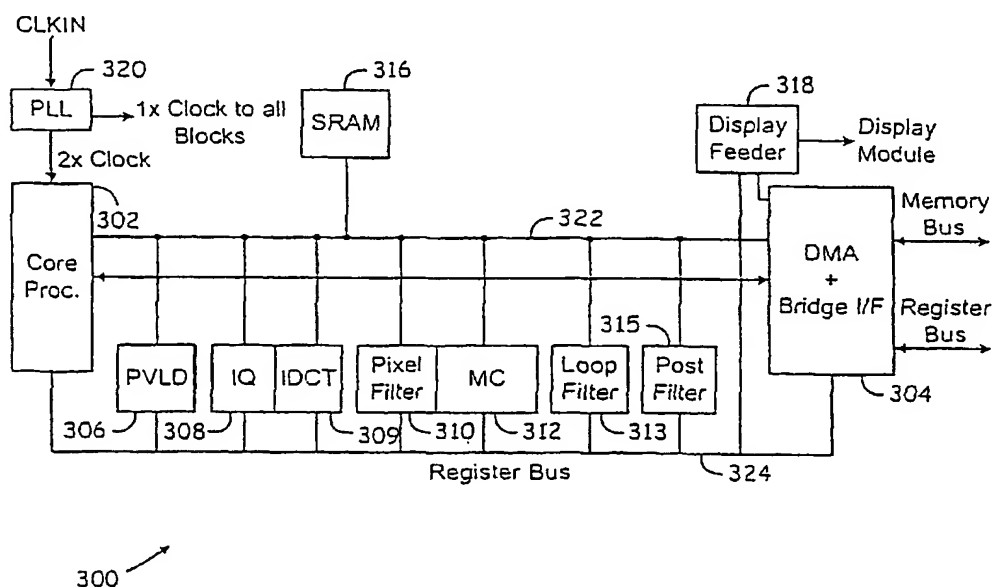
Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:
18 December 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: VIDEO DECODING SYSTEM



(57) Abstract: Video decoding system (Fig. 3) having a programmable variable-length decoding accelerator. The system includes a decoder processor and a variable-length decoding accelerator. The variable-length decoding accelerator is coupled to the decoder processor and performs variable-length decoding operations on variable-length code in the video data stream. The variable-length decoding accelerator is capable of decoding variable-length code according to any of a plurality of decoding methods. In one embodiment, the variable-length decoder includes a plurality of code tables stored in memory and a code table selection register that is programmable to dictate which of the plurality of code tables is to be utilized to decode variable-length code.

WO 03/085494 A3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/09896**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(7) : H03M 7/40

US CL : 341/67

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 341/67, 58, 59, 63

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
NONEElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EAST**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 6,339,386 B1 (CHO) 15 January 2002 (15.01.2002), see entire document.	1-24

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"G" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

29 AUGUST 2003

Date of mailing of the international search report

14 NOV 2003Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

BRIAN YOUNG

Telephone No. (703) 308-1621